CUK-H05920-3-KP129

THE JOURNAL OF PHILOSOPHY

3

Calcutta University Library
ALIPORE CAMPUS

VOLUME LXXXIX, NUMBER 1
JANUARY 1992

page
5 *Structure and Nature* W. V. Quine
10 *The Utility of Pleasure is a Pain for Decision Theory* Anna Kusser and Wolfgang Spohn
30 *Allowing for Understandings* Frederic Schick

COMMENTS AND CRITICISM
42 *Supervenience/Determination a Two-way Street? Yes, But One of the Ways Is the* Wrong Way! Geoffrey Hellman

BOOK REVIEWS
48 W. V. Quine: *Pursuit of Truth* Michael Williams
51 Peter Hylton: *Russell, Idealism, and the Emergence of Analytic Philosophy* Thomas Baldwin

56 MEMO TO AUTHORS

(8)

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S Copyright Law This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER. Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF
PHILOSOPHY

Calcutta University Library
ALIPORE CAMPUS

VOLUME LXXXIX
JANUARY–DECEMBER 1992

Edited by
Bernard Berofsky, Arthur C. Danto
Kent Greenawalt, Charles Larmore
Isaac Levi, Mary Mothersill

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J. E. Woodbridge and J. McKeen Cattell. In 1906 Wendell T. Bush became associated with the JOURNAL as co-editor. In 1923, the JOURNAL was incorporated in the State of New York under its present name. Since its founding it has been published from Columbia University and printed by the Lancaster Press in Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY

100
J826

K-P-129

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers. The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use. Accordingly, the *Journal* authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $.05 per page/per copy is paid directly to us. *Each copy must include a notice of copyright.* We will turn over to the authors any substantial payments (over $30.00, an administrative fee) obtained from all photocopying. Inquiries about photocopying may be sent to the Managing Editor.

© Copyright 1992 by the Journal of Philosophy, Inc.
ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY
VOLUME LXXXIX, NO. 1, JANUARY 1992

Calcutta University
ALIPORE
CAMPUS

STRUCTURE AND NATURE

We are familiar with three adequate but incompatible ways of modeling number theory in set theory, or the theory of classes; and there are infinitely many more. We bandy our numbers without caring which classes we are bandying from among this wealth of alternatives. We are just content that we are operating somewhere within the ontology of classes to which we have committed ourselves anyway for other purposes. This situation, long since remarked upon by Paul Benacerraf[1] and others, receives formal expression in Ramsey sentences.[2] Such a sentence merely affirms that *there are* classes obeying the desired laws—arithmetical laws in this case—and then proceeds to apply such unspecified classes to whatever arithmetical business we had in mind.

That is the structuralist treatment of number. It is just a way of eliminating an idle question—"What is number?"—and a gratuitous decision among indifferent alternatives. Ontological economy is not its point; the same old classes are presupposed.

David Lewis[3] has ventured the next step: a structuralist treatment of classes themselves. Now just as structuralism in number theory depended on the various ways of reducing numbers to classes, so structuralism in class theory depends on reducing classes in turn to—what? For Lewis, as for most of us, I should like to think, there are only classes and concrete individuals. So his structuralism is indeed ontologically significant. It professes a reduction of classes to individuals—hence out and out nominalism.

Arithmetical structuralism, as expressed in Ramsey sentences, depended on there being classes in which the required arithmetical structures could be defined and realized, though we were freed of choosing among alternatives. Similarly for Lewis's structuralism of

---

[1] "Mathematical Truth," this JOURNAL, LXX, 19 (November 8, 1973). 661–79

[2] Frank Ramsey, *The Foundations of Mathematics* (New York: Routledge, 1931), pp. 212–36.

[3] *Parts of Classes* (New York: Blackwell, 1991).

0022-362X/92/8901/5–9

classes; it depends, for its truth, upon there being individuals in which the desired classial structures can be defined and realized. Thus he needs a theory of individuals.

What he uses is mereology, or the theory of part and whole, with atoms. By ingenious constructions he carries the day, contingently on there being enough atoms, and granted George Boolos's[4] view of second-order logic as just a classless logic of plurals. If there are not enough atoms, then the higher reaches of set theory turn out meaningful but false. How much of axiomatic set theory is true comes to depend on the size of the concrete universe: on how many atoms there are. If there are only finitely many, then so much the worse for classical set theory even at its weakest. Truth in mathematics comes to depend on the richness of nature. Lewis appreciates all this.

Uncertain as to the richness of nature, we can still pursue higher set theory in a conjectural spirit, finding out how it would be if nature were up to it. In any event, Lewis's construction renders the whole of set theory *meaningful* from his nominalistic point of view, however much of it is thereby rendered true and how much false. And, ironically enough, the scientific utility of applied mathematics is independent of all that.

Structuralism for classes, hence for all abstract objects, is undeniably congenial. They are things that are known anyway only by their structural role in cognitive discourse; never by ostension. By ostension and extrapolation we do learn what individuals qualify as members of the class of cats, but that class itself is just an abstract entity to which each cat bears a cryptic epsilon relation. Your class of cats and mine can still be different things, and your membership relation and mine can be different, if it makes any sense to say so, though we see eye to eye on every cat. Such is the appeal of structuralism regarding abstract objects.

Much though I admire Lewis's reduction, however, it is not for me. My own line is a yet more sweeping structuralism, applying to concrete and abstract objects indiscriminately. I base it, paradoxical as this may seem, on a naturalistic approach to epistemology. Natural science tells us that our ongoing cognitive access to the world around us is limited to meager channels. There is the triggering of our sensory receptors by the impact of molecules and light rays. Also there is the difference in muscular effort sensed in walking up or down hill. What more? Even the notion of a cat, let alone a class or number, is a human artifact, rooted in innate predisposition and cultural tradition. The very notion of an object at all, concrete or abstract, is a human contribution, a feature of our inherited apparatus for organizing the amorphous welter of neural input.

[4] "Nominalist Platonism," *Philosophical Review*, XCIV, 3 (July 1985): 327–44.

Language, the vehicle of science, is linked to our neural input by neural mechanisms of association or conditioning. Early in our learning of language we learn to associate certain expressions with certain ranges of neural input. I call these expressions *observation sentences*. Grammatically, some are indeed sentences, e.g., 'It's raining', and some are nouns or adjectives, e.g., 'Cat' or 'Milk' or 'White'. To begin with, they are just expressions associated holophrastically with ranges of neural input. They are as if to say, with William James, "Hello, thingumbob again"—except that even 'thingumbob' hints more of objectual reference than I could wish. The observation sentence names nothing, to begin with, neither neural input nor external object. But it is learned from adults who have learned the ways of reference, so it is indeed an expression, such as 'It's raining' or 'Cat' or 'Milk' or 'White', that is destined for eventual integration into a system of objectual reference on the child's own part as he matures. This growing system of terms and purported objects is our science of nature, our organization of the welter of neural input.

Reification and implication are the key principles by which that organizing proceeds. We master implication in the course of learning to use the logical particles, such as 'not', 'and', 'or', 'if', 'every', 'some'. For instance, our learning to use 'not' and 'or' consists in part in learning to assent to '*q*' whenever we have assented to 'not *p*' and '*p* or *q*'. Our learning of 'every' consists in part in learning to assent to '*a* is a *G*' once we have assented to '*a* is an *F*' and 'Every *F* is a *G*'.

It is in implications of this last sort that reification makes its contribution, in our having reified the object *a* and the other *F*s and *G*s. The reification of bodies comes in stages in one's acquisition of language, each successive stage being more clearly and emphatically an affirmation of existence. The last stage is where the body is recognized as identical over time, despite long absences and interim modifications. Such reification presupposes an elaborate schematism of space, time, and conjectural hidden careers or trajectories on the part of causally interacting bodies. Such identifications across time are a major factor in knitting implications across the growing fabric of scientific hypotheses.

At more sophisticated stages in the development of language and science, implications are enhanced by positing further objects, no longer observable; thus subvisible particles, also numbers and other classes.

The implications thus forged are what relate our evolving scientific theory to our heterogeneous neural input and hence ultimately to the external world that our scientific theory purports to describe.

This fabric of interlocking sentences hooks on to the neural input at the observation sentences. Some of these, we saw, were learned by primitive conditioning to ranges of neural input. Many further observation sentences, more sophisticated ones, are scientific sentences retroactively associated with observation through the intervention of theory; thus the chemist sees at a glance that there was copper in the solution, and perceives at a sniff that sulphur dioxide escaped. Now the way implication links theory to neural input is that a bundle of scientific hypotheses is found to imply that the fulfillment of one of two given observation sentences should always ensure fulfillment of the other. This dictates an experiment: he brings about fulfillment of the one and checks for the other. Failing the other, the bundle of hypotheses is discredited; one or another of them must be revoked.

Such, in schematic caricature, is surely the evidential relation of scientific theory to the triggering of our neuroreceptors by the external world. I have ventured more detail elsewhere,[5] and much more wants venturing. Perhaps a fair bit of science might be found on fuller analysis to be beyond the reach of evidence altogether. But this sketch will suffice for what I want to say about structuralism.

The point I now want to make is one that over the years I have repeatedly made in terms of what I call *proxy functions*. The point is that if we transform the range of objects of our science in any one-to-one fashion, by reinterpreting our terms and predicates as applying to the new objects instead of the old ones, the entire evidential support of our science will remain undisturbed. The reason is twofold. First, implication hinges only on logical structure and is independent of what the objects, the values of the variables, may be. Second, the association of observation sentences with ranges of neural input is holophrastic. It is independent of reifications, independent of whatever objects the observation sentences or their parts may be taken to refer to as terms.

The conclusion is that there can be no evidence for one ontology as over against another, so long anyway as we can express a one-to-one correlation between them. Save the structure and you save all. Certainly we are dependent on a familiar ontology of middle-sized bodies for the *inception* of reification, on the part both of the individual and of the race; but once we have an ontology, we can change it with impunity. For abstract objects this is unsurprising, and quite in the spirit of Ramsey, Lewis, and Benacerraf. For familiar bodies it is less intuitive. But we must bear in mind that an observation sentence, say, 'That's a rabbit', retains all its original visual associations; it is only the purported object that is gratuitously displaced.

---

[5] *The Roots of Reference* (Peru, IL: Open Court, 1974); and *Pursuit of Truth* (Cambridge: Harvard, 1990).

This global ontological structuralism may seem abruptly at odds with realism, let alone naturalism. It would seem even to undermine the ground on which I rested it: my talk of impacts of light rays and molecules on nerve endings. Are these rays, molecules, and nerve endings themselves not disqualified now as mere figments of an empty structure?

Naturalism itself is what saves the situation. Naturalism looks only to natural science, however fallible, for an account of what there is and what what there is does. Science ventures its tentative answers in man-made concepts, perforce, couched in man-made language, but we can ask no better. The very notion of object, or of one and many, is indeed as parochially human as the parts of speech; to ask what reality is *really* like, however, apart from human categories, is self-stultifying. It is like asking how long the Nile really is, apart from parochial matters of miles or meters. Positivists were right in branding such metaphysics as meaningless.

But early positivists were wrong if and when they concluded that the world is not really composed of atoms or whatever. The world is as natural science says it is, insofar as natural science is right; and our judgment as to whether it is right, tentative always, is answerable to the experimental testing of predictions.

We saw that reference could be wildly reinterpreted without violence to evidence. We see now that that is just part of a wider picture. Presumably yet more extravagant departures, resistant even to sentence-by-sentence interpretation into our own science, could conform equally well to all possible observations. If we were to encounter such a case and somehow recognize it as such, we might even proceed to master it and then switch back and forth for richer perspectives on reality. But naturalism would still counsel us that reality is to be grasped only through a man-made conceptual scheme, albeit any of various.

My global structuralism should not, therefore, be seen as a structuralist ontology. To see it thus would be to rise above naturalism and revert to the sin of transcendental metaphysics. My tentative ontology continues to consist of quarks and their compounds, also classes of such things, classes of such classes, and so on, pending evidence to the contrary. My global structuralism is a naturalistic thesis about the mundane human activity, within our world of quarks, of devising theories of quarks and the like in the light of physical impacts on our physical surfaces.

W. V. QUINE

Harvard University

THE UTILITY OF PLEASURE IS A PAIN FOR DECISION THEORY*

EVERYBODY agrees that we have intrinsic desires. Obviously, we have desires; many of them are derivative; so, in a kind of analogy to a deductive theory that divides its provable statements into axioms and theorems, some of them must be basic.

There is much less agreement about what our intrinsic desires are or should be. Still, no one doubts that we intrinsically desire, among other things, happiness, pleasure, and the absence of pain, certain qualities of our sensations, feelings, moods, and psychological states in general. We desire these things, but without having, or feeling the need to have, a justification for doing so.

Desires come in various strengths, and, together with more or less firm beliefs, they guide rational action. The formal counterparts of these concepts in decision theory are utility and probability. The distinction between intrinsic and extrinsic desires is also reflected in decision theory, where each model contains a utility function that is basic (at least relative to that model) and an expected utility function that is derived from the basic utility function and the subjective probability function of the model.[1] On the whole, decision theory certainly provides the best formal explication of our intuitive reasoning about desires and beliefs in practical deliberation.

A person's intrinsically desired psychological states should thus be in the domain of her utility function. But then, we claim, decision theory fails. More specifically, we shall defend two theses:

(a) If a decision situation exhibits a certain causal structure, then

---

* We gratefully acknowledge the opportunities to discuss this paper at the Freie Universität/Berlin, the University of Tilburg, and the University of California/Irvine. In particular, we are very much indebted to Jack Birner, David Gauthier, Bert Hamminga, Dan Hausman, Isaac Levi, Gregory Kavka, Karel Lambert, Alan Nelson, Rainer Trapp, and Ernst Tugendhat. This work has been partially supported by the Deutsche Forschungsgemeinschaft, Grant No. Sp 279/2-1.

[1] In the classical theory of L. J. Savage, *The Foundations of Statistics* (New York: Wiley, 1954, 2nd ed. 1972), consequences have basic or absolute utilities, and actions get expected utilities. In the standard framework of decision trees [e.g., H. Raiffa, *Decision Analysis* (Reading, MA: Addison-Wesley, 1968), ch. 2] one may assume a basic utility either for each node or only for each path of the tree. By analyzing the tree by backward induction, one then assigns to each node an expected utility with respect to the subtree starting at that node. In the somewhat different theory of R. C. Jeffrey, *The Logic of Decision* (Chicago: University Press, 1965, 2nd ed. 1983), finally, basic utilities or nonprobabilistic values are assigned only to complete consistent novels or possible worlds, and propositions consisting of more than one possible world have expected utilities.

0022-362X/92/8901/10–29

decision theory is in trouble, because the derivation of expected utilities fails.

(b) This causal structure in fact obtains in a specific, but very common kind of situation, namely, when the intrinsically evaluated psychological states are in the domain of the utility function.

Section I explains thesis (a), section II defends thesis (b). Since the trouble we have concocted may seem to rest somehow on a misunderstanding, section III addresses this suspicion. We do not have a solution to the problem presented below and, indeed, we are not at all sure in what such a solution would consist. Nevertheless, in the final section some constructive conclusions will be adduced.

We shall talk about practical deliberation and decision theory only in an informal or semiformal way, because we shall discuss a general problem that is not produced by any one of the existing formal versions of decision theory. Our considerations are most easily formalized in terms of decision trees or decision-flow diagrams, however, as they are presented, in an exemplary way, in Raiffa (*op. cit.*). Thus, this theory should be used as a formal guideline, if required, and even though our informal intentions are broader, our claims about formal decision theory refer only to it.

It will be apparent that the problem is but a variant of Joseph Butler's[2] criticism of hedonism. Thus, in a sense, the point of our paper is that modern theorizing about practical deliberation has not dealt seriously with Butler's criticism.

I. A CAUSAL ASSUMPTION IN DECISION THEORY

Thesis (a) claims that decision theory presupposes a hidden assumption about the causal structure of decision situations.[3] To uncover it, let us look at how we proceed in practical deliberation.

The task is to find an action that optimally promotes our aims. For this purpose, we list our aims, i.e., the propositions we intrinsically desire to be true, and the acts open to us; each aim has a utility and, for each act, a probability of turning out true conditional on that act. Since each act thereby gets an expected utility, we know which acts have maximal expected utility, i.e., optimally promote our aims.

The calculation of the expected utilities of acts seems to proceed in one step. But usually it does not; usually, deliberation refers to a richer causal picture of the world that draws on the various ways and means by which our acts may influence our aims. This is reflected in a slightly fuller description of deliberation.

---

[2] *Fifteen Sermons Preached at the Rolls Chapel*, 1729, in W. R. Matthews, ed., *Butler's Sermons and Dissertation on Virtue* (London: Bell, 1949), Preface §§ 29–31, and Sermon XI, §§ 1–10. Cf. also H. Sidgwick, *The Methods of Ethics* (New York: Macmillan, 7th ed. 1962), who deals with this criticism to some extent in book I, ch, iv, and in book II, ch. iii, § 5.

[3] There is, in fact, at least one other assumption of this kind. Newcomb's problem has given rise to differing versions of decision theory which converge only when some special conditions are assumed to hold. Cf., e.g., the papers in R. Campbell and L. Sowden, eds., *Paradoxes of Rationality and Cooperation* (Vancouver: British Columbia UP, 1985).

The task is to determine a total utility for each act and indeed, as we shall immediately see, for each proposition in consideration. A very natural assumption—which cannot be adopted as a definition because, as will soon emerge, it is false—is that the total utility $U^t(A)$ of a proposition A is simply the sum of its intrinsic utility $U^i(A)$ and its extrinsic utility $U^e(A)$.[4] Note that a proposition may have both: for some people, money not only makes for a comfortable life, but also acquires an intrinsic magic; driving is fun, but one also runs a risk of getting killed by it.

So deliberation starts from the agent's intrinsic utilities of all the propositions in consideration (many of them will be zero, of course). Then, by a kind of recursive procedure, the extrinsic and total utilities of each proposition are determined. The extrinsic utility of a proposition A is an expected utility; it is the weighted sum of the total utilities of the possible direct causal consequences of A, the weights being the subjective probabilities of these consequences conditional on A. By rolling his causal picture backward from its end points, the agent can thus successively calculate an extrinsic utility and, with the assumption just mentioned, a total utility for each proposition, until he has reached the starting points of his causal picture, i.e., his acts. He thereby finally determines which acts have maximal total utility.

However this sketch of practical deliberation is precisely formalized,[5] its essential point is the conception of desires or utilities and the derivation of them it embodies. There are only intrinsic and extrinsic utilities and sums of them; the intrinsic ones are underived, the extrinsic ones derived, and the derivation proceeds backward from supposed effects and their conditional probabilities and total utilities to supposed causes and their extrinsic utilities.

---

[4] Of course, each utility function always is that of some person at some time, even though we do not make this explicit in the notation

[5] The sketch well agrees with common versions of decision theory. The simple one-step procedure is embodied in Savage's initial representation of decision situations in terms of states of the worlds, acts, and consequences (ch. 2). The more detailed multistep procedure is formalized in decision trees and their analysis by backward induction (see, e.g., Raiffa, ch. 2) which is, by the way, already contained in Savage's more elaborate theory of small worlds (sect. 5.5); our considerations are aligned with this standard formalization. The causal character of practical deliberation is, of course, accounted for in all of causal decision theory which, admittedly, is an elaboration of Savage's theory; cf., e.g., Brian Skyrms, *Pragmatics and Empiricism* (New Haven: Yale, 1984), ch 4.

Talking thus of derivation may suggest that we conceive of decision theory as a syntactic theory about how rationally to compute utilities (and probabilities). This is not so. Decision theory should rather be conceived as stating laws about how utilities and probabilities coexist in rational complexes of graded beliefs and desires. What we mean, then, by using the more graphic phrase that some part of such a complex is derivable from other parts is that the former is uniquely determined by the latter according to the decision-theoretic laws.

Now, if practical deliberation is to work as described, some assumptions of causal well-orderedness have to be satisfied. One natural assumption is that the agent's complex of beliefs and desires entering his practical deliberation is a sufficient cause of his actual behavior. This assumption is, per definition, true of all rational agents.[6] And it is presupposed by practical deliberation in the sense that the weaker the impact of the deliberation on actual behavior, the emptier the deliberation itself. The anomalies that arise with violations of this assumption are well-discussed under the heading "weakness of will."[7]

Another assumption, and the one with which we are here concerned, is that the output of each step of deliberation is not already needed as input of that or an earlier step; otherwise, deliberation would obviously be caught in a circle. The output of deliberation consists in extrinsic and total utilities. How could they possibly be needed as input? When and only when the having of these extrinsic and total utilities is causally relevant to those effects from the utilities of which these extrinsic and total utilities are to be derived.

Let us look at a simple abstract instantiation of this seemingly weird possibility. Suppose that the agent's total utilities of B and non-B are already determined and that A is the only proposition considered by her to be directly causally relevant to B and non-B. Then the derivation of the extrinsic and the total utility of A is straightforward:

(1) $U^e(A) = U^t(B)P(B|A) + U^t(\text{non-}B)P(\text{non-}B|A)$

(2) $U^t(A) = U^i(A) + U^e(A)$

Now grant, for the sake of argument, that in the agent's view B causally depends also on the total utility she actually assigns to A. In this case, the probability $P(B|A)$ in (1) will not do; the deliberation must rather use the probabilities $P(B|A \text{ and } U^t(A) = x)$,[8] for any possible value x of $U^t(A)$. But then the decision-theoretic account of practical deliberation obviously gets circular; in order to derive the value of $U^t(A)$ according to the modified (1), one should know from which value of $U^t(A)$ to proceed, which according to (2), however, depends on the value $U^t(A)$ takes. Hence, for this account to work we have to assume the absence of such a vicious causal dependence.

---

[6] At least for the causalists in action theory who hold that practical reasons may cause actions and indeed do so for rational agents.

[7] Cf., e.g., Donald Davidson, *Essays on Actions and Events* (New York: Oxford, 1980), essay 2; David Pears, *Motivated Irrationality* (New York: Oxford, 1986); and Kusser, *Dimensionen der Kritik von Wünschen* (Frankfurt: Athenäum, 1989), sect 3.1 and 4.2.

We will claim, of course, that this dependence is neither weird nor impossible. But first the argument just given requires four comments on the abstract level.

First, the argument implicitly contains the assumption that what is causally dependent in the agent's view is also stochastically dependent according to her subjective probabilities. We believe this assumption to be defensible; but here is not the place to argue this intricate point.[9]

Second, the argument rests on the claim that the probabilities $P(B|A \text{ and } U^t(A) = x)$ must be used because $P(B|A)$ is unavailable. But this seems false. Why not make $P(B|A)$ available by computing the sum (or the integral) of all $P(B|A \text{ and } U^t(A) = x) \times P(U^t(A) = x|A)$ for all x? The idea of having a probability distribution for one's own total utility of A (this is what $P(U^t(A) = x|A)$ amounts to) looks a bit strange, however. More importantly, the idea is self-defeating, because that distribution would allow one to calculate first $P(B|A)$ and then, by (1) and (2), $U^t(A)$ and thus to gain certainty about $U^t(A)$. Of course, this self-defeat just reflects the circularity of the situation imagined.

The only way to escape this self-defeat would be to start for sure from such a value of $U^t(A)$ that will be confirmed by these calculations. This is the solution we shall suggest and more fully explain in section IV below. But note that this cannot be called a way of deriving $U^t(A)$ because there need not be exactly one such self-confirming value of $U^t(A)$. And even if there is exactly one such value, the procedure results in a determination of $U^t(A)$ that is very different from the one given by (1) and (2) in the absence of the vicious causal dependence. Indeed, in order to grasp fully how drastically the standard decision-theoretic picture is changed by adopting that solution, we first need to understand how that circular situation might come about (see section II) and why one cannot get rid of it within that standard picture (see section III).

Third, there is another tacit assumption in (1) and (2), namely,

[8] Here, '$U^t(A) = x$' stands for the proposition that the agent's total utility of A is x.

[9] For our view on the relation between causality and probability, cf. Spohn, *Grundlagen der Entscheidungstheorie* (Kronberg/Ts.: Scriptor, 1978), sect. 3.3 and 5.1; and "Direct and Indirect Causes," *Topoi*, IX (1990): 125–45.

that the intrinsic utility of A is independent of all the intrinsic utilities entering into the extrinsic utility of A. But this assumption may be dropped. If we allow for utility dependencies, we have to suppose the total utility of A-and-B and of A-and-non-B to be already given (either because the total utilities of these propositions are identical to their intrinsic ones, or because some steps of backward induction have already been performed). Working now one step backward in the decision tree, we obtain the total utility of A alone according to the formula:

$$(1') \quad U^t(A) = U^t(A \text{ and } B)P(B \mid A) + U^t(A \text{ and non-}B)P(\text{non-}B \mid A)$$

But grant again that B causally depends also on the total utility the agent actually assigns to A. Then (1′) gets into the very same trouble as do (1) and (2), for the very same reasons. Therefore, we gain some perspicuity and lose nothing by sticking to the mentioned assumption of utility independence; our problem about deriving extrinsic and total utilities hides in the probabilities and not in possible utility dependencies.

Finally, we have to be a bit more careful in specifying which causal assumption is required for (1) and (2) to work properly without circularity; for one might argue that the vicious causal structure is not in the least surprising. Trivially, the utility function one has is causally relevant to the act one performs[10] and thus also to everything affected by this act. But in our abstract example, for instance, this fact alone would not force practical deliberation to use the probabilities $P(B \mid A$ and $U^t(A) = x)$ instead of $P(B \mid A)$. If the causal influence of the actual value of $U^t(A)$ on B is screened off by A,[11] then these probabilities are the same for all x, and we need consider only $P(B \mid A)$. Thus, deliberation gets circular in this example only if the influence of the value of $U^t(A)$ on B is not screened off by A.

The causal assumption decision theory presupposes is therefore this: if the values of the extrinsic and total utilities of the propositions $A_1, \ldots, A_m$ are at all causally relevant to the propositions $B_1, \ldots, B_n$ from the utilities of which the utilities of $A_1, \ldots, A_m$ are to be derived, then this influence is screened off by $A_1, \ldots, A_m$ themselves. In each case in which this assumption is violated, the decision-theoretic derivation of expected or, in our terms, total utilities is caught in a vicious circle.

This consideration also makes it clear why this assumption is usually satisfied and why it may well have gone unnoticed. Our beliefs and desires certainly have a tremendous influence on the external world, but this influence is exclusively mediated and thus screened off by our acts.[12] Hence, the derivation of expected or total utilities of acts works smoothly as long as the propositions from the utilities of which the derivation starts refer only to the external world; and this is certainly true of most applications actually made of decision theory.

---

[10] This is guaranteed by the above-mentioned assumption about the efficacy of practical deliberation.

[11] This means that, for all x, B is probabilistically independent of the proposition $U^t(A) = x$ conditional on A as well as on non-A

II. A VIOLATION OF THIS ASSUMPTION

Thesis (b) claims that this strange kind of causal situation which jumbles practical deliberation has a very common instantiation. The previous paragraph shows where to look for it; we have to consider our own inner states, in which way they are desired by us and in which way they are caused.

There are certainly many inner states having extrinsic utility for us; a noticeable example is given by beliefs that often have only instrumental value, e.g., for better informed decisions. But, more importantly, we also find a wide variety of sensations, emotions, moods, mentalities, basic frames of mind, and other inner states having intrinsic utility for us;[13] their desirability is of great practical impact and is not derived from other desires in any recognizable way. These states are classified by the occasions that produce them, by the ways in which they are expressed, by their very rich inner phenomenology, of course, and even by their desirability (unpleasant feelings, e.g., have to be undesired). Intrinsically desired inner states will be called *satisfactive states.*

How are satisfactive states caused? This is often opaque, as we all well know. But some general statements seem unassailable. Naturally, our satisfactive states are in the permanent causal grip of the external world. And naturally, the influence of the external world on us also depends on our inner condition it meets. For example, one usually enjoys an invitation with friends, but the more so, the less is one's stress. The crucial question is: What are the relevant parts of that inner condition? And the crucial answer is: among other things, how much we desire the external situations. That is, the inner state a person gets into is produced by the external situation one experiences and by one's prior inner state that includes her prior desire for that external situation often as a relevant part.

It is not difficult to find examples. Two men are watching Super

---

[12] This is an exaggeration, of course; our beliefs and desires are expressed not only in intentional action, but also in unintentional behavior (this fact is most beneficial to our transparency). But as long as we consider ourselves only as agents, as decision theory does and as we here do, we may stick to this exaggeration.

[13] This is not to exclude that these states may have extrinsic utility besides; but nothing of what follows turns on this possibility.

Bowl XXII on television. The Broncos have just scored the first touchdown. One man exults, the other gets nervous. Why? Because the one wants the Broncos to win, the other the Redskins; thus a touchdown by the Broncos has positive utility for the one and negative utility for the other.

Two years ago, a woman got pregnant, and she was miserable. Now she is pregnant again and very happy. Why? Because the first time she did not want to have a baby, but now she wants it. This is a causal explanation. First, there was the desire to have a baby, then she got pregnant, and now she feels happy; but if the desire had been absent, the pregnancy might not have made her happy, as was the case two years ago.

Prior utility and posterior satisfaction need not parallel each other, however. All kinds of relations are possible. The importance of an aim may make me pursue it so grimly that in the end I cannot enjoy the fruits of my efforts. Your badly wanting some kind of thing you are deprived of may work either way: you may not get enough and end up in total frustration, or you may humbly enjoy small amounts of it. There are things or situations that satisfy one only when one does not try to bring them about, or even when one does not desire them.[14]

All these examples exhibit the causal structure that the previous section argued to be problematic:

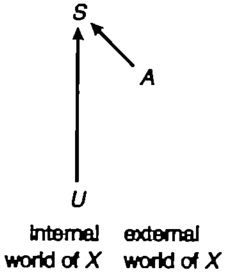


Figure 1

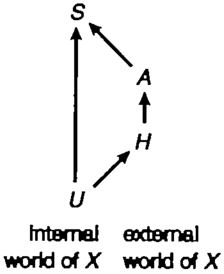


Figure 2

S: a satisfactive state of the person X

A: an external state of affairs

H: an act of X

U: X's inner state of having a specific total utility function that has S, A, and H among its arguments

→ means: is causally relevant to

Figure 1 shows the bare essentials of this structure: they are realized in the football example (with A being the fact that the Broncos have just scored and S the exultation or, respectively, the depression). There is one causal chain running from the external fact A (via television, etc.) to the inner state S. And there is a second causal chain running from U to S which never leaves X's internal world. Its details are not shown in the figure; but it is this chain which provides the psychological basis for how *X* takes up the outside happenings, and *U* contributes to this chain. There is no mention of any acts being contemplated in the example or in figure 1. But the essential point is already contained in this case, namely, that there is a second causal chain giving rise to the derivational circle unfolded in the previous section; hence, the extrinsic and thus the total utility of *A* cannot be derived from the intrinsic utility of *S*.

---

[14] Cf. Jon Elster, *Sour Grapes* (New York: Cambridge, 1983), sect. II.2, for a discussion of this kind of paradoxical situations.

Figure 2 explicitly considers an act and fits the pregnancy example (with *A* being the woman's state of pregnancy, *S* her happiness, and *H*, e.g., the act of stopping the use of a contraceptive). Now, there are indeed two causal chains running from *U* to *S*, the first via *H* and *A*, and the second as before. As stated in the previous section, the first chain poses no problems because the influence of *U* mediated by this chain is screened off by *H*; therefore, the extrinsic utility of *H*, e.g., can be derived from the total utility of *A*. It is again the second chain not screened off by the external world which prevents the total utility of *A*, and thus that of *H*, from being derivable from that of *S*.

Intuitively, deliberation does not get off of the ground at all if the only desire it is based on is the desire to be happy. If you are trying to decide which profession to choose, for instance, you cannot arrive at a conclusion simply by trying to figure out how satisfying the various options would be—not because the decision would be so complex, but because the options as such do not have a definite satisfactoriness for you. Rather, you first need to have at least a rough idea of how much you want to belong to the various professions; only then can you reasonably ask how much they would satisfy you. This is essentially Butler's criticism of hedonism over 250 years ago:

> The very idea of an interested pursuit necessarily presupposes particular passions or appetites, since the very idea of interest or happiness consists in this, that an appetite or affection enjoys its object. It is not because we love ourselves that we find delight in such and such objects, but because we have particular affections towards them. Take away these affections, and you leave self-love absolutely nothing at all to employ itself about; no end or object for it to pursue, excepting only that of avoiding pain (*op. cit.*, Preface, §31).

We propose theses (a) and (b) as a translation of this into modern terms. If plausible, they point to an important incompleteness in decision theory. The gap does not open in each case in which our satisfactive states are part of the deliberation. Sometimes, the deliberational circle is side-stepped by, say, overriding moral considerations; sometimes the influence of the causal chain within the person's internal world is negligible; and it would be most interesting to look at these exceptions more carefully. But in many ordinary cases the circle remains.

III. HOW NOT TO DENY THE CIRCLE

Although the present puzzle for practical deliberation may seem easily invented, there seems to be no easy way out. The following five suggestions, at least, will not do.

(A) It may be thought that we have relied too heavily on the intuitive causal picture of practical deliberation and its formal representation in decision trees, and that the circle vanishes in some other suitable formalization of decision theory. But this idea does not seem plausible if the candidate formalization is one of the current versions of so-called causal decision theory because they have the same roots as those to which we have referred. One might refer instead to some kind of so-called evidential decision theory; but we are skeptical whether this helps. In any case, the burden of proof does not seem to lie with us.

(B) One might think that we have simply offered an old refutation of hedonism in a new disguise, that only the most stubborn still believe in hedonism, and that we are beating a dead horse. This, however, would be a misunderstanding; our problem is a more general one. We do not merely criticize the hedonistic claim that the only things intrinsically evaluated are satisfactive states. Rather, our concern is with the usual decision-theoretic explanation of what it means to derive extrinsic from intrinsic values; and for that purpose we needed only to assume that some of the things intrinsically evaluated are inner states of ourselves. Thus, everybody who adheres to this explanation is in trouble, not just the hedonist.

Still, one might say that the problem emerged within a hedonistic venture, i.e., in the attempt to derive extrinsic utilities of external states of affairs from intrinsic utilities of satisfactive states. So why not simply abandon this attempt? But, surely, this is not yet an answer. The utilities of external states of affairs somehow relate to the intrinsic utilities of satisfactive states, even though this relation is not decision-theoretic derivation. We should not deny this relation, but try to account for it.

(C) Another suggestion is that the puzzle somehow is generated by an ambiguous use of the term 'utility'. In particular, one might think that, in the terms of figure 1, the utility of A that is causally relevant to S—let us call it $U^c(A)$, for the moment—is not the same as that utility of A that is to be derived from that of S'. So what is $U^c(A)$? Is it really $U^t(A)$, as we have said?

One might suggest that $U^c(A)$ is a kind of evaluation different from the kind used in practical deliberation. But it need not be, of course. We do not want to deny that propositions (or other things) are evaluated by us in other, e.g., aesthetic respects and that these other kinds of evaluation may be causally relevant to our satisfactive states. But we insist that the decision-theoretic kind of evaluation used in practical deliberation has this causal relevance, too; and all the examples given confirm that claim. (Even in the football example, the desires of the television watchers actually are of no practical relevance because they cannot do anything for their favorite teams; but they would, if they could.)

One might still object that there are now three alternatives for $U^c(A)$: $U^c(A)$ may be taken as the intrinsic, the extrinsic, or the total utility of A. $U^c(A)$ might indeed be $U^i(A)$; but then there is no deliberational circle because $U^i(A)$ is underived and available in advance. $U^c(A)$ cannot be $U^e(A)$; if it were and if we are right about the circle, there would not exist any derivable extrinsic utility of A. Finally, it seems that $U^c(A)$ cannot be $U^t(A)$ either; $U^t(A)$ was assumed to be the sum of $U^i(A)$ and $U^e(A)$, and if $U^e(A)$ does not exist, $U^t(A)$ does not either. Thus, the objection concludes, it seems to be an utter mystery which kind of utility should generate the circle.

There is a formal answer. In anticipation of this objection, we have emphasized in section I that it is only an assumption, not a definition, that $U^t(A)$ is the sum of $U^i(A)$ and $U^e(A)$. If this assumption is dropped—as is forced by the circle—the last step of the objection fails, and the identification of $U^c(A)$ with $U^t(A)$ can be preserved.

Our examples provide also an intuitive answer. The woman's happiness *is* influenced by how strongly she desires a baby; your satisfaction *is* influenced by how strongly you desire the profession you choose. In these cases, the desire affecting later satisfaction clearly is not (or not only) an intrinsic desire. In the woman's case, it is her total desire for the baby which springs from many sources: the intrinsic desire for the baby, the disutility of the various costs, the desire to please her husband or even her parents, etc., and, of course, also the desire for the satisfaction she derives from the baby. In the case of the profession, it is easily conceivable that the profession is in no way intrinsically desired, but only meant to satisfy, to yield earnings, etc.

Hence, the right conclusion from the objection is that there is something wrong not with our intuitive description of the circular cases, but with the usual classification of utilities into intrinsic, extrinsic, and total ones. The examples point to a kind of total utility which, though not intrinsic, cannot be decision theoretically derived in the presence of the circle. What we need is a more adequate classification of utilities.

(D) One might still feel, however, that the puzzle arises because it is not really clear what is meant by utility. The puzzle heavily relied on the causal wedge driven between utility and satisfaction; and this distinction may appear illegitimate. Indeed, many have tended somehow to equate the one with the other. We have to discuss separately the two ways of construing pronomial reference here; but it should be clear what we are up to: if you (1) reduce utility to satisfaction, then you do not do justice to utility; and if you (2) reduce satisfaction to utility, then you do not do justice to satisfaction. There is a way (3) of bridging the difference between utility and satisfaction with which we sympathize; but, as we will see, it does not resolve our deliberational circle either. Let us look at (1), (2), and (3) a bit more thoroughly.

(D-1) The agent, as depicted in decision theory, is quite a strange individual. He is always very busy calculating utilities and trying to maximize expected utility, and he seems to be done when having done so. But does he ever get anything? What is his ultimate pay-off? In the simple case where money takes the role of utility, the answer is easy: what he gets is money. In the real case where utility must not be equated with money, it is natural to give a similar answer: what he gets is utility. Indeed, what sense is there in trying to maximize the expectation of a quantity one does not get? So far, however, utility is only a name for the agent's ultimate pay-off. Now we know, of course, what his ultimate pay-off is: it is how he feels and how he is; in our terminology, it is his satisfactive states measured on some scale as conceived in the old calculus of pleasure. *This* is what utility is.

That is, briefly, the consideration that leads naturally to reducing utility to satisfaction; and many have been tempted by it.[15] But it does not yet give a complete account of utility. It only says that the agent's future utility pay-off consists in his future satisfactive states. But how are we to understand his present utility function? There seems to be only one way: the present utilities of his possible future satisfactive states are just the degrees of satisfaction they would give him were they realized; and the present utilities of other states of affairs represent expected satisfaction, i.e., they are expected utili-

---

[15] This reduction is characteristic of classical utilitarianism, but it may also be found in modern welfare theory. Cf. Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, J. H. Burns and H. L. A. Hart, eds. (London. Athlone, 1970), ch. I; J. S Mill, *Utilitarianism*, in *Collected Works*, F. E. L. Priestley, ed. (Toronto: University Press, 1969), vol x, pp. 203–59; Sidgwick, book II; and A. C. Pigou, *The Economics of Welfare* (New York: Macmillan, 1920), pt. I, ch. II. See also A. K. Sen, "Plural Utility," *Proceedings of the Aristotelian Society*, LXXXI (1980/81): 193–215; and J. Griffin, *Well-Being* (New York: Oxford, 1986), ch. 1.

K-P-129

ties representing the supposed conduciveness of these states to future satisfaction.

This conception is nevertheless inadequate. Its picture of practical deliberation and rational agency contains only the agent's possible future satisfactions and his present beliefs about them, while totally dispensing with the agent's desires or wants. We take it as obvious that such a picture is fundamentally distorted.

One may try to improve this picture by locating the agent's desires in it, i.e., by defining them within it. This, however, looks unpromising from the beginning, because desires are widely taken as a basic kind of propositional attitude not reducible to other propositional attitudes.[16] Moreover, the only way of locating them seems to be to say that the agent's present utilities, which have already been defined, simultaneously represent his present, more or less intense desires. But this move turns hedonism into an a priori truth; it stipulates in effect that the agent's desires are intrinsically concerned *only* with his future satisfactive states. Moreover, the need of fitting this stipulation to the intuitive notion of desire entails a traditional problem of hedonism, namely, that of having to relate each and every desire to the experience of satisfaction and each and every preference to a difference in (the expectation of) the experience of satisfaction. Again, all this seems distorted.

Thus, the crucial point is, in a nutshell, that present desires may not be identified with present expectations about future satisfactions and that present utilities must represent the former and not the latter in order to play the role for practical deliberation and rational agency which they actually play in decision theory.[17]

Finally, it must be noticed that this conception of utilities does not make the deliberational circle disappear; it only transforms it. Our claim (b) then amounts to the claim that the agent's present expectation of future satisfaction is causally relevant to future satisfaction; and it is supported by precisely the same examples. It says, in simpler nonprobabilistic terms, that the agent's present belief that this state of affairs will produce that future satisfaction is causally relevant to whether it actually is so. This kind of belief is, so to speak, self-verifying (or in some other way causally affects its own truth). And it generates an analogous kind of deliberational circle: to know which future satisfaction you should expect, you should know which future satisfaction you do expect. Hence, this conception of utilities is not only misguided, it is also unhelpful, and we should resist it.

---

[16] See also D. Lewis, "Desire as Belief," *Mind*, XCVII (1988): 323–32, for a general argument why desire is not reducible to belief.

[17] This is more thoroughly argued in Kusser, sect 3.3.

(D-2) In fact, the history of the concept of utility in this century is shaped by such resistance, as is reflected, e.g., in the so-called theory of revealed preferences. Most thinkers, in particular economists, nowadays conceive utilities as representing various desires of varying strength as they are revealed by the preferences and actions of rational persons. But then the question of the relation between utilities and satisfaction re-emerges.

There has been a long-standing tendency to reverse the answer just discussed, i.e., to say that satisfaction is whatever we have when our desires come true and thus to turn satisfaction into a purely formal notion.[18] Probably no one has put this so bluntly because it sounds so absurd. But the tendency to conceive satisfaction (or pleasure or happiness) in this way has ample historical precedent. For example, you find Immanuel Kant[19] saying:

> Glückseligkeit ist der Zustand eines vernünftigen Wesens in der Welt, dem es im Ganzen seiner Existenz alles nach Wunsch und Willen geht (*ibid.*, p. 224).

And John Rawls[20] saying:

> . . . we are happy when our rational plans are going well, our more important aims being fulfilled, and we are with reason quite sure that our good fortune will continue (*ibid.*, p. 548).

Nevertheless, we should resist any such tendency. We have got what we wanted too often without being satisfied; and the phenomenology of what we have called satisfactive states is too vivid and too independent. So, no conceptual assimilation of satisfaction to realization of desires can be successful.

There are, of course, reasons for this tendency. One reason certainly is that it is at least the normal case that the realization of our desires satisfies us; but not every case is normal. A stronger reason is that in each case we have to assume that the realization of our desires will satisfy us. If we believe that some of our desires will not satisfy us and if we still stick to these desires, then we either have overriding, say, moral reasons for doing so, or we are plain foolish. The situation resembles G. E. Moore's paradox of belief: we have to believe that what we believe is true; but others may know better, and we may learn to know better, too. Likewise, we have to believe that what we desire will satisfy us. But others may know better, and we may learn to know better, too. Such reasons may have furthered this tendency, but they do not justify it.

---

[18] This is nicely paralleled by a purely formal notion of pay-off utility as something you get or have whenever a desire of yours gets fulfilled. Cf Griffin, ch. 1, and Sen, sect. 4.

[19] *Kritik der praktischen Vernunft*, 1788.

[20] *A Theory of Justice* (Cambridge: Harvard, 1971).

(D-3) Once we acknowledge that the relation between satisfactions and desires or utilities, i.e., the grounds for rational action, is contingent, the deliberational circle stands. There is some support for the view that, in some sense, our genuine or *true* desires are those which actually satisfy us and that we ideally have such desires.[21] We share this view, and we will expand on it below. But it has no impact on the deliberational circle. This circle is generated by the relation between actual desires or utilities and satisfactions and does not vanish by considering how desires ideally are.

(E) You might still feel uneasy about the theses leading to the circle. Perhaps you are tempted by the following objection. "Crudely put, your thesis (b) says that very often a person enjoys something because she wants it, and then you somehow extract a problem from this. But if you ask this person why she wants that thing, it would be natural for her to reply: 'Because I enjoy it.' And there is nothing wrong with that reply—except that you cannot accept both the reply and your thesis (b); you cannot have it both ways."

One can easily be misled by this consideration; indeed we have been. There is, however, no incoherency at all; the two 'because's in that objection simply have two different meanings. When we say that a person enjoys something because she wants it, we state one of the many causes of her enjoyment. When she says that she wants it because she enjoys it, she expresses one of the many practical reasons for her want.

It is almost as simple as that, but not quite; two problems hide in this answer:

(E-1) We have just admitted that the person's belief that she will enjoy something is a practical reason for her desiring it. But what is a practical reason? Is *A* not a practical reason for *B* just in case *A* is essentially used in a practical deliberation supporting *B*, i.e., in a decision-theoretic derivation leading to *B*? The problem is that the last sentence contains in fact two different explanations; the first is vague and correct, the second precise and incorrect. The conclusion we have to face is that there are two kinds of practical reasons: the decision-theoretic kind and some other kind. This other kind is instantiated by the belief that one will enjoy something, but it still lacks an explanation.

---

[21] Cf. Bertrand Russell, *The Analysis of Mind* (London: Allen & Unwin, 1921) ch. III, esp. pp. 72ff.; J. C. B. Gosling, *Pleasure and Desire: The Case for Hedonism Reviewed* (New York: Oxford, 1969), ch. 7; Ernst Tugendhat, *Probleme der Ethik* (Stuttgart: Reclam, 1984), pp. 33–56, esp. pp. 45f.; and A. Tversky and D. Kahneman, "The Framing of Decisions and the Psychology of Choice," in Elster, ed., *Rational Choice* (New York: University Press, 1986), p. 138.

It would not be correct to require of a rational person to have true desires; we also do not require of her to have true beliefs. Seeking, not having knowledge is essential to rationality (of course, the second often is the result of the first). Thus, it certainly belongs to rationality to strive after true desires. This sense of rationality may be intended in the above quotations of Kant and Rawls.

(E-2) The causalists in action theory, with whom we have obviously sided, hold that, for a rational person at least, what is a practical reason from the person's inside perspective turns into a cause from our outside perspective. Now, when we admit that the person's belief that she will enjoy something is a practical reason for her for desiring it (and think that she is rational), we also have to say that her belief causes her desire. And since we keep claiming that her desire is causally relevant to her actual enjoyment (and think that causal relevance is transitive), we have to conclude that her belief that she will enjoy something is causally relevant to its truth. Should we really be prepared to accept this conclusion?

IV. HOW TO TAKE THE CIRCLE SERIOUSLY

It seems that the circle stands unshaken, but each attempt to avoid it opens questions. If the circle is important, it presumably calls for a theory. We have none; but we want to suggest some constructive observations and some positive conclusions.

We have hinted at the rudiments of a formal solution of our problem already in section I; let us spell them out by looking again at formulas (1) and (2). In these equations, the relation between A and B was assumed to be regular, so to speak. In the circular case, however, which can only obtain when B represents a future satisfactive state, the causal beliefs of the agent are expressed not by $P(B|A)$, but rather by $P(B|A \text{ and } U^t(A) = x)$, for any possible value x of $U^t(A)$. Therefore, the former should be replaced in (1) by the latter. Now, if $U^t(A)$ were known to take the value x, the modified (1) and (2) would determine a value y of $U^t(A)$. Thus, if the value of $U^t(A)$ is x, it should be y; and this is satisfiable, only if $x = y$. In other words, only solutions of the equation

$$(3)\quad x = U^i(A) + U^t(B)P(B|A \text{ and } U^t(A) = x) + U^t(\text{non-}B)P(\text{non-}B|A \text{ and } U^t(A) = x).$$

are eligible as values of $U^t(A)$; otherwise, the total utility of A does not match the expected satisfaction A derives from B, given this total utility of A.

The fact that the unknown x appears both at the left hand and the right hand side of (3) nicely reflects our deliberational circle. Note that no extrinsic utility is now assigned to A; when total utility is constrained by (3), it can no longer be assumed to be the sum of intrinsic and extrinsic utility. Equation (3) may have no, exactly one, or, as will be quite common, more than one solution, depending on how the conditional probabilities in (3) vary with x. If the mathematical problem were neatly formalized, one could presumably bring fixed-point theorems of equilibrium theory into action. We have not attempted to do so, for reasons stated below.

Nevertheless, the equilibrium idea embodied in (3) is intuitively very helpful; it informally clarifies six of the questions we have left open.

Four kinds of utility. Equation (3) specifies the relation between the utilities of satisfactive states and those of external states of affairs which was sought in (B) of the previous section. This relation complements the decision-theoretic picture of practical rationality. According to this picture, there are only two kinds of utility: utilities figuring as derivational premises, i.e., the intrinsic ones, and derived utilities, i.e., the extrinsic and total ones. In (3), however, utilities have different roles. It would be inappropriate to say that in (3) $U^t(A)$ is derived from $U^i(B)$ or from $U^t(B)$.[22] This is particularly obvious in the case where (3) has more than one solution. Thus, the intrinsic utilities of satisfactive states like B do not in general function as derivational premises. Correspondingly, the utilities of many external states of affairs like A are neither derived nor intrinsic in the original sense. Rather, the relation between $U^i(B)$, or $U^t(B)$, and $U^t(A)$ in (3) is more appropriately described as one of *control*. The utilities of external states of affairs are controlled by those of satisfactive states in the sense that the former have to be in equilibrium with the latter and have to be changed whenever they are out of equilibrium.[23] Utilities may also be subject to such an equilibrium control—this is the essential lesson to be drawn from our deliberational circle.

In this sense, (3) generates two new kinds of utility: controlling and controlled utilities. This also clears up the muddle in (C) about which kind of utility of A is causally relevant to satisfaction. It is the underived total utility of A—where 'underived' does not mean "intrinsic" in the old sense, and 'total' does not mean "derived." There is nothing paradoxical in this assertion, once the four kinds of utility have been recognized.

Two kinds of practical reasons. What we have just said also answers the question in (E-1) of the previous section. There are indeed two kinds of practical reasons: those functioning as premises in a decision-theoretic derivation and those supporting equilibria like solutions of (3). The desire for satisfaction and the belief that one will enjoy something—or, more generally, utilities and probabilities like the ones on the right hand side of (3)—are reasons of this second kind. Thus, if asked why one wants something, there is nothing wrong with the answer: "Because I enjoy it." What would in general be wrong is the assumption that this want is simply derived from the desire to feel joy.

---

[22] $U^i(B)$ and $U^t(B)$ will usually be equal if B is a satisfactive state, as it is in (3).

[23] Indeed, the relation is less one-sided than the word 'control' suggests. One may re-establish equilibrium also by re-evaluating satisfactive states. Thus, when you experience that the fulfillment of a certain desire does not satisfy you, you may also try to learn to draw satisfaction from it, e.g., by redirecting your awareness, by depreciating alternative satisfactions, etc. Indeed, this method of accommodating pleasures to desires (instead of desires to pleasures) may often be observed. It should not be dismissed as irrational.

The causalists' claim that practical reasons viewed from inside are causes viewed from outside has to be specified accordingly. It seems clear that this claim can only be maintained for the decision-theoretic kind of practical reasons for which it was intended. The practical reasons of the equilibristic kind at best cause one to stick to the desire supported by them, but not to have it in the first place. We thereby can avoid the undesired conclusion of (E-2); but the next point looks into the matter more carefully.

The status of the belief that the realization of a desire will satisfy. What we have just said means that this belief accompanies rather than precedes the desire. Indeed, it is a necessary companion of the desire. Unless one has other reasons for a desire, one would be irrational to have the desire and lack the belief that its realization will satisfy. This or rather the probabilistic counterpart thereof is what (3) asserts; and this agrees well with what we have observed in (D-2).

What about the suspicion lurking in (E-2) and also at the end of (D-1) that this belief is not a normal empirical belief, but rather one of the self-verifying kind? We have to be careful here concerning the precise content of this belief. It may be (i): the realization of a desired proposition, say, A, will satisfy. Or it may be (ii): the realization of the proposition A as a desired one will satisfy. Anyone who believes (i) will probably also believe (ii); and anyone who believes (ii) and is conscious of his desire for A will certainly believe (i). Still there is a subtle difference. If one's causal belief is expressed by (i) and one overlooks or denies that one's desire for A is causally relevant to how satisfying A is, then one can maintain the received picture—at the cost that the belief in (i) is in fact causally affecting its truth. If, however, one's causal belief is expressed by (ii), one has to give up the received picture with respect to the utility of A—with the benefit that the belief in (ii) is a normal empirical belief. Only this second alternative can be sustained. The first is self-defeating; as soon as one realizes that one's belief in (i) is affecting its own truth, one can no longer stick to it.

Changes of the belief that the realization of a desire will satisfy. When we said that this belief necessarily accompanies the desire, this did not mean that this belief would merely be a by-product of the desire. On the contrary, under reading (ii) it is a normal belief to which normal epistemology applies. It may be grounded or ungrounded by evidence; it may be false; one may learn that it is false; and, to put it so generally as to cover also the probabilistic case, one may change it on the ground of new evidence according to general rules of rational belief change. But which consequence has such a change for the desires? Again, we have to distinguish two cases.

The normal case is that belief change only calls for a recalculation of the expected utilities. This may even be true of changes of the beliefs about future satisfaction. For instance, I might decide in a restaurant to try an exotic dish that I have never tasted before. It turns out that I do not especially like it, and I accordingly revise my belief about the satisfying efficacy of that kind of dish and thus its expected utility. This case allows such a simple description, however, only when our circle is absent, i.e., when one's desire for a dish does not affect how much one likes its taste.

The case is very different when our deliberational circle is present. The change of the beliefs about future satisfaction, i.e., of the probabilities involved in (3), then results in a predicament: the controlling desire for satisfaction and the controlled desires for other things are out of equilibrium and can no longer be maintained rationally. Somehow, they must be changed; somewhere, a new equilibrium must be found; but there is no hint about how to change them and where to seek it. Everyone will know the experience of running on empty—e.g., when the thrill of rock 'n' roll fades or when one's love is worn out—and the often difficult process of redirecting one's desires.

Equation (3) as a model of deliberation. Equations (1) and (2) provided rules of practical deliberation; they told us how to derive extrinsic and maximize total utilities. Equation (3), which replaces (1) and (2) in the cases discussed, might thus also be taken as offering a model of deliberation, namely, this one: solve (3); if it has no solution, that is too bad (though this may be ruled out by suitable continuity assumptions); if it has a solution, fix $U^{i}(A)$ to be the maximal solution—because among all the eligible values of $U^{i}(A)$ this is the one which gives you the highest expected satisfaction! If this were a feasible model of deliberation, then it should be helpful to work out the mathematics. But we doubt this very much, for two reasons.

First, the relevant probabilities seem scarcely to be available; for knowing them means having an answer to all the questions of the following kind: "If I should desire A so and so strongly, how likely is it that the realization of A will satisfy me to such and such a degree?" But we simply do not know and thus cannot grade many of the possible satisfactions. How does it feel to be a free climber? How satisfying would it be? We have not the slightest idea. And it is even more difficult to take into account the influence of the desire. How satisfying would it be to be a free climber, if one really wanted to be one? Thus, we can plausibly answer these questions only for the kinds of satisfactions we have experienced and for the kinds of desires we have pursued; all other probabilities would be artificial guesswork.

Secondly, the formation of aims and ends, i.e., of underived desires is quite obscure; the idea that one could adopt, by sheer will, so to speak, the desires that such deliberation recommends, looks a bit strange. Thus, it seems doubtful whether such deliberation could really be as effective as it should be in order to have a point.

Rational and true desires. If (3) should not be understood as a model of deliberation, what does it then accomplish? It provides a model of *learning*—at least in the weak sense that it says that desires in disequilibrium should rationally change and that this change should rationally move to some new equilibrium.

Indeed, (3) establishes an extended notion of rational desire. In the received picture, there was no way at all of assessing intrinsic desires or utilities as rational or irrational; this had a precise meaning only for extrinsic desires. In our amended picture, however, (3) provides a standard of rationality also for controlled as well as for controlling desires (and thereby opens a possibility of grasping the notion of rational feelings). This is tantamount to saying that (3) provides extended means of criticizing the desires of a person, not according to moral or otherwise external standards, but strictly by that person's own measures.

Insofar as the beliefs that the realization of desires will satisfy may be true, or insofar as the probabilities in (3) may be the objective ones, (3) finally gives some meaning to the notion of *true* desires: our true desires are those which are in a maximal equilibrium given true beliefs. We rationally search for true desires; and since the causation of our satisfactions keeps changing, we have to search again and again. This is what we in fact do. We do not gain practical help from (3), but perhaps some theoretical insight into what we are seeking.

ANNA KUSSER

Freie Universität/Berlin

WOLFGANG SPOHN

Universität Bielefeld

ALLOWING FOR UNDERSTANDINGS

LET me begin with some history. In days gone by, the annual American Philosophical Association meetings were held at colleges and universities, and there was always a banquet, at which the president gave his address. One year at the end of the 1950s, the APA met at Columbia. Being then young and just marginal members, my friends and I were placed at the banquet in a far corner of the room. How marginal we were came out soon enough, while the others were eating, when we saw that we hadn't been served. It turned out the kitchen had run out of food. We were least likely to make a fuss and so had been picked for doing without.

I made the suggestion that we leave and get burgers and come back for the talk, but this found no takers. One person sitting at that table was loudly and firmly opposed to it. He said he had already spent $5.00 for dinner, and wouldn't spend a penny more. I said that, if we left, we would each get a burger and pay just $2.00, while if we stayed we would starve. He replied that that was ridiculous, that if he did what I proposed, he would in the end have paid $7.00 and only got a burger for it. He said he would rather starve. The others agreed, and so we stayed and ate celery.

The reader may need some assuring: there is a philosophical issue in this. That issue is, I think, deep and important, and it connects with many others. Still, it has been almost wholly ignored. The reason may be that seeing it clearly calls for giving up certain convictions that are usually held without question. The issue can be fully grasped only in a proper framework of thought. I will discuss the framework first and leave the problem of the burgers for later.

Consider this experiment by Daniel Kahneman and Amos Tversky.[1] People were put in two situations. Each person in situation I was told to assume he had just got 1,000 (in Israeli pounds). He was then asked to choose between (A) getting another 500 for sure and (B) an even chance of getting another 1,000 or getting nothing more. In situation II, each subject was told to assume that he had been given *2,000* and was asked to choose between (C) *losing* 500 and (D) an even chance of losing 1,000 or losing nothing. The majority of the people in I chose (A), and the majority of those in II chose (D). If they had been in both situations (first in one and then the other), many might have chosen both (A) and (D).

---

[1] "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, XLVII (1979): 263–91.

0022-362X/92/8901/30–41

There seems to be something wrong, for suppose that the outcomes are reported in terms of the chooser's final assets in them. Since getting 1,000 and then 500 more comes to the same as getting 2,000 and losing 500, etc., the two situations both reduce to III, to having to choose between (E) winding up with 1,500 and (F) an even chance of winding up with 2,000 or 1,000. It seems that, if anyone chose (A) and (D), he faced the same issue twice, and the first time chose one option, the second time the other.

How do Kahneman and Tversky account for the likely (A) plus (D) choices? They suggest that people, in choosing, focus not on total final positions but on the gains or losses their actions might generate for them: "The apparent neglect of a bonus that was common to both options [the 1,000 in one case and 2,000 in the other] . . . implies that the carriers of value or utility are changes of wealth, rather than final asset positions that include current wealth" (*ibid.*, p. 273). Hence the priority assigned to the setups in I and II over that in III. They also suggest that the value curve for money is roughly S-shaped, that it goes through the origin, and is generally concave for gains and generally convex for losses. This last implies that where, as in situation I, a person must choose either a gain for sure or a chance of gaining and these options have the same money expectation, he is risk-averse, and that where, as in situation II, he must choose either a loss for sure or a chance of losing with the same money expectation, he is risk-inclined: hence the choices of both (A) and (D).

Notice the similarity to the burger problem. In that, I held that, if we went out to eat, the outcome would be a burger for $2.00, a clearly positive gain, that the $5.00 we spent for the banquet was now long gone and irrelevant. My friend—let me call him Sidney—held that the local context matters, and that the past can't be kept out of that. I reasoned as did the (A)-and-(D) choosers, whom Kahneman and Tversky thought foolish. That is, I reasoned as if I took "the carriers of value . . . [to be] changes of wealth. . . ." Sidney sided with Kahneman and Tversky; he thought in terms of "final asset positions." I saw a gain in store if we left and only a small chance of dinner if we stayed. Thinking of the gains made me averse to the risk, and so I suggested we leave. Sidney noted that our final position if we left would be negative (counting what we already had spent, he saw it as a loss) and that we still had a chance of eating (of breaking even) if we stayed. This inclined him to take the risk and so he argued for staying.

A second account is also possible of how all these choices were made, and Kahneman and Tversky brush shoulders with it. They preface their report of their experiment by saying: "We now show how choices may be altered by varying the representation of outcomes" (*ibid.*). Here it isn't a matter of outcomes versus final asset positions but of different presentations of the outcomes, of different reports given of them. And indeed the (A)-and-(D) choosers, unlike the hypothetical (E) and (F) choosers, built the gains and losses they expected directly into their descriptions of the options. The authors make no allowance for such differences in their explanatory "Prospect Theory," nor do they think they can be used to rescue these people's rationality, for they hold that a sensitivity to presentations marks a failing of its own. A person who changes as the presentations change (where he knows the presentations are coreportive) violates what they call "invariance." Thus he departs from what they consider essential to common sense.[2]

In several papers and in a book just published, I have taken the opposite view. The way people put the facts to themselves I have there called their *understandings* of them, how these people *see* those facts. I have held that taking note of understandings is essential not only to explanations of actions but also to any judgment of whether some choice or action was rational. My case for the agent's understandings being essential to an explanation of his conduct rests on some general theses. The first is that desires and utilities focus on propositions, that the "bearers of value" are not bare physical facts but facts as the agent reports them. Also that any fact can be differently reported or propositionalized by the agent—differently *understood* or *seen* by him. Also that propositions reporting the same are sometimes assigned different utilities by a person even where he knows they are coreportive, and that the way a person acts reflects the utilities (and probabilities) he assigns.

Here our topic isn't explanation but judgment, the judgment of rationality, but much the same can be said. My contention that a person's understandings bear on judgments of the rationality of his conduct stems from my wanting a logic of action that is more open, more indulgent than the usual. And I want such a logic because I think that the (A)-and-(D) choosers made the right choices, given their understandings, though Kahneman and Tversky, given *their* understandings, would have been right to choose differently. Also that my stand on the burgers was sound, given my understandings, and that Sidney's was too, given his. In other studies,[3] I have worked out such a logic, one that allows for each of the courses in (several of) Kahneman and Tversky's experiments, and also in Allais' problem, in Newcomb-like problems, and in various others—a logic that approves different courses for people who see the same things differently.

---

[2] See their "The Framing of Decisions and the Psychology of Choice," *Science*, CCXI (1981). 453–8, and their "Rational Choice and the Framing of Decisions," in Robin M. Hogarth and Melvin W. Reder, eds., *Rational Choice: The Contrast between Economics and Psychology* (Chicago: University Press, 1987), pp. 67–94.

I won't go into these matters now. What I want to do instead is to attend to the likely critics, to consider the sort of objections that might be brought against my whole project. These fall under three separate headings: ideological, formal, and visceral. Some comment is needed on each.

The first sort will not keep us long, though the issue here is central. I want to arrange for a logic of action that allows for people's understandings. This means that the logic of action I want has to be *intensional*. More strictly, it means that its theory of value has to be intensional, that the theory may not fault someone's setting different utilities on propositions he thinks coreportive.[4] If propositions that were thought coreportive had to be given the same utility, it couldn't matter which propositions reported the facts as the agent understood them. A person's understandings couldn't matter on their own. Many writers hold that they *oughtn't* to matter: propositions that we think coreportive ought to be valued alike by us. These writers hold that a proper logic has to be *ex*tensional.

The issue is central, but on this level, nothing much can be done with it; everyone just beats his own drum. Here are Kahneman and Tversky again, likening people's judgments of value to their judgments of height. "Veridical perception requires that the perceived relative height of two neighboring mountains, say, should not reverse with changes of vantage point. Similarly, rational choice requires that the preference between options should not reverse with changes of frame."[5]

Kahneman and Tversky speak of preferences only, but they would say the same of utilities: rationality requires that the utilities we set do not change with our understandings. We face some confusion here, for the analogy of the mountains is poor. The measurement of height is real-object oriented; our "vantage points" or "framings" of an object are irrelevant to the height that it has. By contrast, the measurement of utility is oriented to propositions—we evaluate propositions—and propositions *are* framings. So the thesis of intensionality is not what the mountains analogy suggests. It isn't that different framings make for different valuations of the facts (the facts themselves don't get valued), but that different framings of facts can have different values set on them, ***and this even where these framings are framings of what is thought the same fact.***

---

[3] The logic is put most fully in my *Understanding Action* (New York: Cambridge, 1991).

[4] My use of "intensionality" is not the usual, but it retains the core idea of the nonsubstitutivity of certain related propositions; see *Understanding Action*, pp 95–101

[5] "The Framing of Decisions and the Psychology of Choice," p. 453.

Let me beat that drum a bit more. The self-same fact can be variously framed: it can be variously propositionalized. Thus my basic idea is not that the way we frame or understand things affects the utilities we set. Rather, the idea is that our understandings select the propositions the utilities of which will matter, ***and might even settle on propositions whose utilities are different from those of others we think are coreportive with them.*** I hold that, in picking those propositions whose utilities will then figure for us, our understandings affect our choices. I also hold that this needn't be faulted—that we needn't insist on "invariance."

Again, that is the stand I take, but ought it to be endorsed? Ought a logic to allow for it? Isaac Levi,[6] speaking for the negative, says that "[o]ne of the advantages of [a logic that doesn't] is that its prescriptions are robust under transformations of the way possible outcomes are represented" (*ibid.*, p. 342). This begs the question before us: Is it indeed an advantage? My claim is that it is not, no more so than it would be an advantage if the prescriptions of a logic were robust under transformations of the agent's probabilities. The robustness of a logic is its ignoring (or abstracting away from) certain kinds of variables, and such ignoring isn't always a virtue. In particular, it isn't one here.

We turn now to some matters of logic. An ideological resistance to my not ignoring understandings need not stand in opposition alone. It may be supported by the contention that thinking like mine meets with formal problems, that it runs afoul of basic principles of utility theory. Here there are serious difficulties.

Let $u(h \& k)$ be the utility a person sets on *h-and-k*. Let $u(h, k)$ be the utility of h conditional on k, the agent's degree of disposition to want h if, in addition to all else he believed, he now believed k. (This covers the case in which k is an item the agent already believes.) A familiar principle of conditional utilities is

(1) $u(h, k) = u(h \& k)$

Suppose that the agent thinks that h and n are coreportive. Then he believes that h is true if and only if n is. Putting $h\text{-}iff\text{-}n$ for k in (1), we get

---

[6] "A Note on Newcombmania," this JOURNAL, LXXIX, 6 (June 1982): 337–42.

$$u(h,\ h\text{-}iff\text{-}n) = u(h\ \&\ h\text{-}iff\text{-}n) = u(h\ \&\ n)$$

With h and n interchanged, we get

$$u(n,\ n\text{-}iff\text{-}h) = u(n\ \&\ n\text{-}iff\text{-}h) = u(n\ \&\ h) = u(h\ \&\ n)$$

It follows that

$$u(h,\ h\text{-}iff\text{-}n) = u(n,\ n\text{-}iff\text{-}h)$$

But since the agent believes $h\text{-}iff\text{-}n$ and so also $n\text{-}iff\text{-}h$, this comes to

$$u(h) = u(n)$$

In sum, where a person thinks that h and n are coreportive, the utility he sets on h must be the same as that which he sets on n. Or rather, this follows from (1), the principle of conditional utilities. Principle 1 yields extensionality (invariance), just the constraint I want to avoid. I seem to be in some trouble here.

Extensionality is also imposed by the usual principle of expectedness. Let $p(k, h)$ be the probability of k conditional on h. The usual principle is

$$\text{(2)}\quad u(h) = p(k, h)u(h\ \&\ k) + p(m, h)u(h\ \&\ m)$$

where $p(k\text{-}\&\text{-}m, h) = 0$. In this, put $h\text{-}iff\text{-}n$ for k and $not\text{-}(h\text{-}iff\text{-}n)$ for m. Since the agent believes $h\text{-}iff\text{-}n$, $p(h\text{-}iff\text{-}n, h) = 1$, and so $p(not\text{-}(h\text{-}iff\text{-}n), h) = 0$ (both by the axioms of probability). Principle 2 thus says that $u(h) = 1 \times u(h\ \&\ h\text{-}iff\text{-}n) + 0$, which comes to $u(h) = u(h\ \&\ n)$. Doing this over with h and n interchanged, we find that $u(n) = u(n\ \&\ h)$. Here too it must be that $u(h) = u(n)$. Intensionality is ruled out again. (The rulings-out are related, for (1) can be shown to follow from (2).)

I think the message is this. The usual logic does indeed impose extensionality on us. It might be said that is part of its purpose, part of what it is *for*. If we want more freedom here—if we want to free ourselves to set different utilities on propositions we think coreportive—we need a less constraining logic. We have to weaken the usual principles in some suitable way.

I will put that off for a moment, for there is also a second problem. When I argued for getting a burger, I did this because I saw my getting it as *I get a burger for $2.00.* I knew of course I had spent $5.00 already and that the outcome thus might have been put as *I get just a burger, having spent $7.00.* I preferred getting a burger for $2.00 to starving, though I preferred starving (that is, eating celery) to getting just a burger when I had spent $7.00. Let *I get a burger for $2.00* be ***h***, and let *I already spent $5.00* be ***k***. Then *I get a burger, having spent $7.00* is ***h-and-k***. Since I believed ***k***, it follows by (1) that I should have valued ***h*** and ***h-and-k*** alike, should have equivalued *I get a burger for $2.00* and *I get a burger, having spent $7.00*. And so (1) rules out the preferences I had, for those bound me (by transitivity) to preferring ***h*** to ***h-and-k***.

The general point is that our usual logic implies a certain clumping. I will say it *agglutinates:* wherever a person believes any ***k***, he must assign to every ***h*** the utility he assigns to ***h-and-k***. Whether ***k*** enters his understanding of what ***h*** reports doesn't matter. (Whether my already having spent $5.00 entered my understanding of getting the burger doesn't matter.) A person in the first of Kahneman and Tversky's situations should have valued *I gain 500* and *I wind up with 1,500* alike (the latter being equivalent to *I gain 500 and I just got 1,000*). A person in their second situation should have set the same utility on *I lose 500* as on *I wind up with 1,500*. Likewise, a person should set the same utility on *I find a dollar bill* as on *I find a dollar and then one day I die*.

Some people will bite the bullet and say that the utilities *should* be the same, even in the dollar-finding case. (Since we know we are mortal, our reaction to that should color all else.) I think that they *might* be the same, but that our logic should not demand it. So I think that the usual principles need to be weakened here too. Notice that agglutination is more basic than extensionality: only where they agglutinate must we agree that utilities are extensional. If we can block the former, we can be free of the latter.

We want to weaken the principles so as to give due weight to people's understandings. Here we will apply the concept of their *conditional* understandings—their dispositions to understand things if they believed, say, ***k***. Let me say that a proposition that reports a fact the way the agent understands that fact is a *compelling* proposition for that agent (we might have also called it *salient*). We will say ***h*** is ***k-pre-emptive*** for him where ***h*** would be (or would remain) compelling for him if he believed ***k***. We will say ***h*** is ***k-adoptive*** for him where it isn't ***k***-pre-emptive and, moreover, if he believed ***k***, he would understand the fact ***h*** reports as the conjunction ***h-and-k***.

Back now to principles 1 and 2. We will add this proviso to (1), that ***h*** is ***k***-adoptive: (1) will hold for us only with that proviso attached. We will add the proviso to (2) that ***h*** is both ***k***-adoptive and ***m***-adoptive. These provisos are being added to principles that present the utility of ***h*** in terms of the utility of that proposition conjoined with one or more others. They confine the force of these principles to where that (or those) other proposition(s), if believed, would contribute to the agent's understanding of what is reported by h.

With the provisos in place, our logic no longer agglutinates. No problem now with my having set different utilities on *I get a burger for $2.00* and on *I get a burger, having spent $7.00,* for *I get a burger for $2.00* was not *I-already-spent-$5.00*-adoptive. The former was pre-emptive for me; though I knew I had spent $5.00 already, I still saw my getting that burger as *I get a burger for $2.00.* Likewise (I would guess) for the (A)-and-(D) choosers and also for the dollar-bill finder. To the average person, *I find a dollar bill* is *one-day-I-die*-pre-emptive; though he knows he will die one day, he still sees his finding a dollar as just *I find a dollar.*

Extensionality is lifted too. Suppose again that the agent believes that *h-if-and-only-if-k*. The new (1) and (2) here get us $u(h) = u(n)$ only in certain special cases, only where h is (h-*iff*-n)-adoptive and n is (n-*iff*-h)-adoptive. They get us $u(h) = u(n)$ only where the fact reported by h is understood by the agent as h-*and*-n (that is, as h-*and*-(h-*iff*-n)) and the fact reported by n is understood as n-*and*-h. And so we get this only where the agent sees the fact(s) reported by h and by n the same way. That gives us all the sensitivity to understandings we need.

The reader may want a basic principle less saddled with provisos than (2). If so, he might accept

$$(3)\quad u(h) = p(k, h)u(h, k) + p(m, h)u(h, m)$$

which holds given only that $p(k\text{-}\&\text{-}m, h) = 0$. Nothing excessive follows from (3). Where we believe k, we get just $u(h) = u(h, k)$—we are spared agglutination. Where we believe h-*iff*-n, we get just $u(h) = u(h, h\text{-}iff\text{-}n)$ and $u(n) = u(n, n\text{-}iff\text{-}h)$—we are spared agglutination and the consequent extensionality. Before we qualified (1) as we did, (3) and (2) were equivalent, but (3) is now the more general principle: it holds where (2) does not. Where h is k-adoptive and m-adoptive, (3) reduces to (2) via (1). Where it isn't, (2) does not apply but (3) does apply.

For some, this may be too much—here we arrive at the visceral part. Too many headaches, too many provisos, and now a new general principle! Thank you very much, but no! People oughtn't to monkey with logic. The classical principles are simple and clear. They work well enough for our purpose. Where they conflict with judgments we make, it is those judgments which should be changed.

The question is: Do they really work? I don't think that they do. They are, I think, too constraining. They impose extensionality and agglutination. Thus they screen off distinctions that matter, distinctions of how things are seen. That often makes for confusion, and it has left a mess of conundrums in decision theory. This of course has to be argued for, but I won't do that here.[7]

The basic negative gut reaction may be that the above is ad hoc. I noted what I took to be problems and tinkered with things until they were blocked. The critic may say this is too self-indulgent. He may insist that a settled logic has a claim to greater respect. Such a logic ought not to be changed to suit every contrary whim, no more so than we should weaken our ethics to cater to every temptation.

Yes, of course, about whims and temptations, but what is a whim and what is not? And what is *too much* self-indulgence? The golden rule of metaphilosophy is that of reflective equilibrium: "A [principle] is amended if it yields a [judgment] we are unwilling to accept; a [judgment] is rejected if it violates a [principle] we are unwilling to amend."[8] I am unwilling to accept the judgments of extensionality and agglutination. The critic is unwilling to amend the familiar traditional principles of utility. Here we must each of us go his own way, and I have gone by way of attaching provisos regarding understandings.

The reader may make the objection that I haven't said what understandings are. That of course is correct. I haven't said what beliefs are either, or desires or preferences or probabilities or utilities. These are all basic mental states (the probabilities here are *subjective* probabilities), states too basic to be defined in terms of anything else. The nature of beliefs and desires and the rest emerges only in a theory of mind, in what such a theory says about them. In part, it comes out in a theory of reasons, in how that theory says they combine to make for action. The same is true of understandings: their nature too comes out in such a theory. I have said briefly why I believe the explanation of action must take account of understandings. I will here have to leave it at that.[9]

We can now finally speak of the issue the burger problem raises. Should we have left to get burgers? The answer I gave to that was proper, given my understandings (and my beliefs and desires), and Sidney's answer was too, given his. Logic—that is, my modified logic—faults neither him nor me. It endorses both our positions, given our different understandings. But what about those understandings now? Was his more sound, more right than mine, as he implied by calling me ridiculous, or was mine more sound than his? What makes an understanding sound, and what makes an understanding *unsound*? Or is there no such thing as soundness, and every understanding is as good as any other?

---

[7] I argue for it in *Understanding Action*.

[8] Nelson Goodman, *Fact, Fiction, and Forecast*, 4th ed., (Cambridge: Harvard, 1983), p. 64

[9] Again, I put my position fully in my *Understanding Action*

Regarding the issue between Sidney and me, one might perhaps hope to find help in economics. For our issue was that of sunk costs, of whether money already spent should count in deciding between courses still open. Should the counting start from now or from some point in the past? Economists side with me on this question of whether to let bygones be bygones. Here is a typical statement: "[A sunk cost] should play no role (except as a help in forecasting . . .). . . . For any ensuing decision, only the escapable, or 'variable' costs are relevant."[10] But the point is taken for granted; there is no argument offered for it. So it cannot settle the issue against someone (e.g., Sidney) who disagrees.

Let us get beyond Sidney and me, beyond just the issue of sunk costs. Our two different understandings didn't exhaust the possibilities. Someone else may have seen the outcome of leaving the banquet as *we eat at the deli,* or as simply *we eat*—no mention of costs in that. A vegetarian would have seen it as *we resort to eating meat.* All of these propositions would have been *true* had my suggestion been taken. The question is whether any (or all) of them would have been *appropriate,* whether any (or all) would have served to express a proper understanding of the outcome. Or better, the question we have is this, what decides such an issue? Are there standards of proper understanding, and if so, what are they?

The question here is very broad. The understandings a person has enter his reasons for whatever he does. They enter because the components of his reasons—his beliefs, desires, utilities, and the like—focus on propositions reporting the facts *as this person understands them.* Is his understanding in a situation improper? The reason that moves him must then be too: the beliefs and desires and the like he brings up are not then those he should have brought up. So the issue of the propriety of understandings seeps into that of the propriety of reasons and the question is hard to avoid. What standards, if any, apply? What makes an understanding (and thus a reason) proper? When are we seeing things right?

We look in vain for an answer to the classics. The few who remark on understandings at all try to shake off the question of their rightness. Aristotle does it by definition. For him, an agent's minor premise reports his "grasp" or "understanding" of some option, but also, understandings *must* be right; they wouldn't be understandings if they were wrong: "Understanding is identical with goodness of understanding, men of understanding with men of good understanding."[11] Since not all people are in this elite, we need a word for how those not in it see the facts—'*mis*understand' will do. This leaves us the question of how to distinguish understanding from *mis*understanding, and that is no different from the question on which we are looking for help.

---

[10] Armen A. Alchian, "Cost," *International Encyclopedia of the Social Sciences* (New York· Free Press, 1968); repr. in Harry Townsend, ed., *Price Theory,* 2nd ed (Harmondsworth: Penguin, 1980), p. 232.

Here is Kant on understanding, which he refers to as "judgment": "judgment is a peculiar talent which can be practiced only, and cannot be taught."[12] He also puts it differently, that the special talent is that of making *right* or *proper* judgments. He speaks of judgment as the application of rules—as the grasping or shaping of facts to which our laws (the rules) might apply—and he holds that "the power of *rightly* employing them," of rightly applying them, is a "natural gift" (*ibid.;* emphasis added). An exercise of this natural power yields a right judgment by definition. (Kant is here joining with Aristotle.) The difficult problem is with us still. It is to distinguish what is and what isn't an exercise of this power. When is a person judging correctly? When is he seeing things right? On that we still have no clue.

Among the moderns, one might look to Goodman, who studied a special case of this question (*op. cit.*, pp. 72 ff). A hundred marbles, all of them blue, have been drawn from an urn; the next to be drawn will very likely be blue. All the drawn marbles were also *bleen,* a bleen marble being one that is either drawn before tomorrow and is blue or is drawn after and is green. Parity of reasoning suggests that the next marble is likely to be bleen, but this says that if it is drawn tomorrow it is likely to be green and so *not* likely to be blue. Formally identical inferences from the same evidence put in different ways give us contradictory conclusions. We avoid the contradiction by rejecting the bleen descriptions. But what is our warrant for rejecting them? What makes these descriptions improper?

Goodman develops an answer in terms of the record in prior inferences of the basic words involved (of the words 'blue' and 'bleen'). Whether or not this approach of his works, its focus is very narrow. It deals with a special question only. It has to do just with what descriptions are proper for inferential purposes—not with right seeing in general, but with right seeing in inference only. The general question remains: When is a description, in any context, proper? When are we seeing or understanding things right?

There are thinkers who reject this question, who think it is misguided. They deny that one can speak of the rights and wrongs of understandings except insofar as the judgments involved have to do with some suitability or coherence. They are content to ask how an understanding advances the agent's purposes. Or they may ask how it squares with our own or with social custom, whether that understanding is widely shared, or is the sort that most people have, or the sort accepted by people in some group to which the agent belongs. Right and wrong have no strict application: people whose seeings or understandings differ live in (have built) different worlds, and no judgment of propriety can be made of one world by those who live in another.[13]

---

[11] *Nicomachean Ethics*, 1143a

[12] *Critique of Pure Reason*, A133, B172.

Similar stands have also been taken regarding desires and even beliefs, that there are no rights or wrongs there. We have the beliefs and desires we have, and if they cohere, no more can be said. There are no external constraints; everyone can suit himself. I resist this libertine view, though I have no good case against it. In the end, it may win by default, but in the meantime I put it off. I hold it is wrong to think the earth flat and wrong to want to push people around, and that these judgments of wrongness can be grounded in some impersonal, externalist way.

So too with understandings. I hold that some seeings are wrong, that it is wrong to see marbles as bleen, even where they are—that even a marble we know to be bleen ought to be seen as blue or as green—and that this is more than mere preference. Also, on the political front, that it is wrong to see affirmative action (hiring) as a sort of discrimination, though in fact that is what it is, and that this is not just my policy bias dressed up to look more impressive. I hold it is wrong to see abortions as killings, though that is what they are, and wrong to see executions for murder as arrangements for getting even. But then I still face the question, what grounds do I have for saying all this? What makes it wrong to see these things so? When is a seeing right and when wrong?

Having now asked this several times, I will let it go.[14] I have no better handle on it than does anyone else. So we have made little progress. We began with some hungry philosophers around a table thirty years ago, and we wind up with empty bowls here again. Still, let me look at the bright side. I fancy that the question we have is a deep and difficult one. It is a question that calls for an answer and a question that has been ignored. So at least we have something to chew on, which is more than we had back then.

FREDERIC SCHICK

Rutgers University

---

[13] This last is Goodman's position; see his *Ways of Worldmaking* (Indianapolis: Hackett, 1978). See also Richard Rorty's writings, esp. his *Contingency, Irony, and Solidarity* (New York: Cambridge, 1989).

[14] I consider some answers and why they fail in *Understanding Action*. Also in my "Under Which Descriptions?" in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (New York: Cambridge, 1982).

COMMENTS AND CRITICISM

SUPERVENIENCE/DETERMINATION A TWO-WAY STREET? YES, BUT ONE OF THE WAYS IS THE *WRONG WAY*!*

IN a recent article in this journal, Richard B. Miller[1] argued that supervenience/determination relations are "metaphysically insignificant" because they hold trivially and symmetrically between domains of interest: e.g., not only does the physical trivially (sic) determine the mental, the moral, the aesthetic, etc., it is equally true that the mental, the moral, the aesthetic, etc., each determines the physical. The appearance to the contrary is an illusion resulting from not granting the gander ("higher-order" domains such as the mental, the moral, etc.) what we grant the goose (the physical domain), namely, a language rich in its capacity to express fine and far-reaching distinctions, via mathematical, relational, and dispositional predicates. Extending similar expressive capacity to the higher-order languages results in the claimed symmetry—hence "metaphysical insignificance"—of supervenience.

These claims are, in my view, without merit, and it is the purpose of this note to indicate why. This will also afford an opportunity to clarify certain aspects of supervenience/determination concepts relevant to their intelligent application, and to underscore how it is that they indeed mark major metaphysical, scientific distinctions and are worth pursuing. The key points are the following three.

(1) The formulation of "physicalist supervenience/determination" Miller works with is tendentious. It says, in effect,

> IPI: Any two [actual] things that differ in higher-order respect *R* differ [somehow] physically.

The bracketed words highlight the two key problems with the formulation. IPI stands for "Identity of Physical Indiscernibles," as it was dubbed by Frank W. Thompson and myself[2] (cf. formulas (2) and (3), *ibid.*, p. 555).

Now, it is easy to trivialize supervenience/determination with such a formulation, for indeed any two things do differ physically, on various constructions of what this means (with appropriate caveats

---

* I am indebted to Jaegwon Kim for helpful correspondence.

[1] "Supervenience Is a Two-way Street," this JOURNAL, LXXXVII, 12 (December 1990): 695–701.

[2] "Physicalism: Ontology, Determination, and Reduction," this JOURNAL, LXXII, 17 (October 2, 1975). 551–64

0022-362X/92/8901/42-7

for quantum field-theoretic entities, whose status as "things" is in any case problematic). For example, of course, (classical) physical things differ at least in their spatiotemporal locations, and location properties are the subject of physics.

Indeed, the supervenience literature contains many different formulations of determination relations (as Miller noted, 696, fn. 4). But all serious formulations have *modal force*.[3] Whatever the specific machinery employed, the aim has been to capture intuitive theses such as,

> *Were* any two objects sufficiently similar in relevant physical respects (taking account of their environments), they *would* also share higher-order properties.

In fact, the original formulation of "physical determination" (fn. 2 above) was explicitly designed to improve upon IPI, which was the weakest formulation of "no difference without a physical difference" considered: it does not even quite imply purely ontological physicalism, and was explicitly regarded as palpably too weak to capture the idea of "determination." The principles articulated as genuine determination principles employed model-theoretic notions as a way of expressing modal force. Also, they were framed globally, building in no limits to the extent of environment that might be relevant. They were different ways of making precise the informal condition:

> Any two global (mathematically standard) models of scientific laws ($\approx$ any two naturally possible worlds) indistinguishable in relevant physical respects, are indistinguishable in higher-order respects as well.

Note that this calls for "interworld comparisons," and it allows us to consider, for example, "counterparts" that share spatiotemporal properties. The fact that any two actual things have different locations becomes entirely irrelevant.

---

[3] For a useful survey, see Paul Teller, "A Poor Man's Guide to Supervenience and Determination," *Southern Journal of Philosophy*, XXII, Supplement (1984): 137–62. See also Thompson and my *op. cit.*, and our "Physicalist Materialism," *Noûs*, XI, 4 (1977). 309–45; Kim, "Concepts of Supervenience," *Philosophy and Phenomenological Research*, XLV (1984): 153–76; and John Haugeland, "Ontological Supervenience," *Southern Journal of Philosophy*, XXII, Supplement (1984): 1–12.

It is true that W. V. Quine—whom Miller cites in addition to the above—can be read as resting with the nonmodal IPI as articulating his "physicalism." But the supervenience/determination literature has been unanimous from the start in going beyond this, and, as indicated, Thompson and I explicitly investigated and rejected the IPI as a determination thesis. It is therefore ironic (to say the least) that Miller should seize on this very formulation as the basis for a critique of supervenience!

As Jaegwon Kim has suggested (in correspondence), however, the fact (if indeed it is one) that any two distinct (concrete) things in any nomologically possible world differ in some physical respect (even mere location) is hardly without metaphysical, conceptual interest. The standard involved here is just the necessitation of IPI above, and was designated "weak supervenience" by Kim (*op. cit.*). Furthermore, the corresponding claims of "identity of moral, or of psychological, indiscernibles" (at least in their necessitated form) seem wholly implausible.

Moreover, modal force must be combined with suitable restrictions on the physical base. Not every relational property, for instance, even if expressed in physical language, counts as a ***relevant physical respect.*** In any real application of supervenience/determination to physicalism or related theses, such restrictions are obviously necessary and have been intended all along. Without them, the slightest difference in the physical ontology of two worlds would generate "physical distinctions" between counterparts (however indiscernible locally). For example, predicates of the form 'coexists with . . .', if admitted as physical, would trivially generate such irrelevant distinctions. Again, the original formulation of "global physical determination" was implicitly based on a list of physical predicates derived from actual scientific practice. Moreover, ***physical locality*** can be used to motivate further severe restrictions.[4] Indeed, once this is recognized, even intraworld comparisons do not automatically lead to trivialization. (For example, absolute location may be deemed irrelevant in accordance with standardly adopted physical symmetry principles; so may minor differences in various physical magnitudes, and so forth.) Taking all this into account, it becomes by no means trivial that, for instance, any two relevantly physically similar persons must have the same mental experiences.[5]

Thus, the first (and, I think, decisive) step in seeing through Miller's "goose-gander" argument is to realize that he has not considered the genuine article: a live, robust goose, not a cooked one, and not a straw one either.

(2) Concerning what the gander can do, Miller claims (700) to have considered "the toughest alleged counterexamples" to such reverse theses as "moral determination of the physical," but in fact far tougher ones than any presented readily suggest themselves. So

---

[4] See my "Determination and Logical Truth," this JOURNAL, LXXXII, 11 (November 1985): 607–16.

[5] As it stands, such a schematic principle is of course vague. Serious study of such a claim can only proceed by looking closely at the relevant sciences—in this case, the neurosciences and psychology—in order to ascertain just what properties are relevant as determinants of what.

long as the examples are restricted to a world with sentient life (not to say intelligent or human life, as in Miller's own examples), a case could perhaps be made that any physical difference "makes a moral difference," if we allow sufficiently far-flung dispositional predicates into the moral "base." Consider, however, two worlds (or world stages) which differ physically but in which all forms of life are necessarily precluded (e.g., very early stages of the cosmos in a big-bang model in which matter as we know it cannot exist). (The physical differences could be minor, or even major for that matter.[6]) What corresponding moral differences can there be between those worlds? By hypothesis, the possibility of moral agents (at any time in those worlds, if we like) is ruled out! It makes no difference to entertain what would be the case, were moral agents present, since the assumption of such agents involves a contradiction (with fixed background theory). Appeals to spiritualism that might occur at this juncture are quite beside the point. We may presuppose *ontological physicalism,* for the sake of argument, since it is a necessary, independent component of physicalism (with much going for it, one might add).[7] Implicitly, Miller must be granting this anyway, for, if spirits are allowed, obviously physicalist supervenience can readily fail, whereas Miller charges that it is instead trivially true. Similarly, we may insist that 'moral' means "moral by human standards" (or something perhaps more refined). The standards of other possible agents—possible spirits, God, what have you—are irrelevant. This much of our current usage must surely be held fixed!

(3) The situation is even worse with regard to Miller's claims on behalf of "phenomenal determination of the physical" (in his penultimate paragraph, 701), where he says, among other things, that "if we make the physical differences fall below the threshold of human sense organs, there are always instruments. When all else fails, we can always postulate superhumanly sensitive perceivers to take up the slack."

To take the latter suggestion first, in physical worlds in which sentience is not even physically possible, dispositional predicates involving "superhuman perceptual capacities" are without effect (again, unless a form of spiritualism is being embraced). We cannot effectively entertain what superperceptual difference a physical difference would make in such worlds, for no worlds are countenanced in which the physical phenomena of these hostile worlds occur together with any possible superperceivers.[8]

---

[6] In my "Realist Principles," *Philosophy of Science,* L, 2 (1983): 227–49, I explicitly considered such cases, in order to undermine the sort of *tu quoque* objection Miller makes, but in connection with phenomenalism, rather than "moralism."

[7] The need for purely ontological (physicalist) principles, formulated independently of determination principles, was stressed in Thompson and my *op. cit.* (1975).

With regard to the appeal to scientific instruments, two remarks are in order. First, they too are, a fortiori, physically impossible in the sorts of worlds (or world stages) under consideration, since, by hypothesis only radiant energy is possible—stable atoms cannot even be formed. So counterfactuals involving that which instruments—in the pertinent sense of *epistemically accessible instruments*—would measure are without effect. Second, it is outright cheating for the phenomenalist to make use of terms referring to physical measuring instruments in the first place (quite independently of the first point). Such terms are part of the (macroscopic) physical vocabulary, and unless they can be phenomenally reconstructed, their use in the present context could at best support merely a claim of determination of the physical by the phenomenal + the macrophysical. But of course the situation is worse, for no one has ever come close to providing a thoroughgoing phenomenalist substitute for such macrophysical terms (one of the salient reasons being the vicious circularity of appealing to other instruments or material background conditions relevant to the perceptions an actually unperceived object "would give rise to," as Wilfrid Sellars, J. J. C. Smart, and others forcefully argued decades ago).

Thus, when properly framed, determination relations are not only nontrivial; they point to deep asymmetries among major alternative metaphysical frameworks that have inspired philosophical inquiry over many centuries.

This is, of course, not to say that there is a unique (minimal) adequate determination base; the prospect of multiple alternative such bases (ones not even mutually intertranslatable) has also been recognized.[9] Nor is it to say that asymmetry always prevails. In fact, the article under consideration here does suggest an amusing way of so stretching the notion of "morally relevant property" that an example of codetermination may be in the offing.

To see how this can arise, consider the following sort of scenario: Dr. Strangelove (pretend in the actual world) has his finger on the

[8] Concerning early-cosmos type examples: if one wishes to consider temporally complete cosmological models, then care must be taken to ensure either that sentience is ruled out for all time (in both worlds considered), or that the early differences remain forever causally and epistemically inaccessible The latter is readily arranged, as early evolution of the cosmos is described, e.g., by Stephen Weinberg in *The First Three Minutes* (New York: Basic, 1977)

Such complexities could be bypassed by choosing even simpler examples, e g , worlds with a very small number of elementary particles and minuscule total energy, so that (given our fixed background theory) all possibility of sentience ever arising is ruled out.

[9] See, for example, the "codetermination" standard of theoretical equivalence in Thompson and my *op. cit.* (1977).

button and resolves to push it just in case (he ascertains that) there is a possible world of type *A* that also has arbitrary physical property *B*, where *A* is possessed by both worlds of any proferred counterexample to "moral determination of the physical" and *B* is the allegedly morally irrelevant physical property distinguishing one of the worlds from the other. (*B* can be as trivial as you like, e.g., a minor density fluctuation, a particular value of momentum of a particle, whatever.) Now, *B* appears to be "morally relevant," for the world that satisfies it also satisfies 'gives Strangelove an excuse to blow up the Earth'! (Strangelove, being an accomplished physicist, may even set about trying—singlehandedly—to deduce whether an *A* & *B* model is a solution to, say, the Einstein equations!) Therefore, allowing in dispositions such as 'could give a (possible, if not actual) Strangelove an excuse to . . .', we can sustain even a modalized version of "no physical difference without a (Strangelovian) moral difference"! What should one say about such a line?

Of course, on balance the worlds of any such example are morally on a par (in any ordinary sense of "equally good or bad"), since Strangelove (or a *Doppelgänger*) could interchange the roles of *B* and not-*B*, etc. Still, one might say such alternative scenarios just correspond to so many more morally relevant differences that *B* makes, whose total effect is to leave *B* and not-*B* "equally good."

One thing to say is that a predicate such as 'gives a (possible) Strangelove an excuse to . . .' at best shows that a given prima facie morally irrelevant physical property ***could become*** morally relevant, not that it ***in fact*** "makes a moral difference." In contrast, the dependence of the psychological, say, on the physical is not merely a potential dependence. It is not merely that a mental event, say, a feeling of revulsion, ***might*** be attendant on neural events constituting, say, a reaction to perception of mass slaughter; it is that it ***is*** attendant on such things, and in lawful ways that science may illuminate.

Another thing to say is (the obvious) that such Strangelovian properties are "morally adventitious": they have nothing to do with intrinsic moral worth. And if it is claimed merely that ***the morally adventitious determines the physical,*** that may be granted as a truism (assuming we are allowed to play the above (Peter) Sellersian game). Indeed, one would not attach great metaphysical significance to ***this*** thesis. We see it as an artifact of the (stipulated) open-endedness of "morally relevant property." This, however, hardly impugns the metaphysical relevance of supervenience/determination. The most that it shows is that, indeed, there are some (metaphysically) uninteresting cases.

GEOFFREY HELLMAN

University of Minnesota

BOOK REVIEWS

Pursuit of Truth. W. V. QUINE. Cambridge: Harvard University Press, 1990. viii, 113 p. Cloth $17.50.

The aims of this little book are, in W. V. Quine's own words, to "update, sum up, and clarify" his "variously intersecting views on cognitive meaning, objective reference, and the grounds of knowledge" (vii). A marvel of compressed exposition, it should appeal to long-time students of Quine, who want to keep abreast of his current thinking, as well as to newcomers seeking an introduction to the views of the most influential American philosopher of his generation.

Though not drastically revisionary, the book shows a general drift toward a softening of some of Quine's more radical theses. The familiar Quinean themes are all in evidence, but frequently played in more muted tones. Take Quine's conception of epistemology: while still holding that there is no first philosophy, firmer than science and serving to "ground" it, Quine retreats from his sometime view that this turns epistemology into a straightforward chapter of psychology. He now maintains that, within the "baffling tangle of relations between our sensory stimulation and our scientific theory of the world, there is a segment that we can gratefully separate out and clarify without pursuing neurology, psychology, psycholinguistics, genetics, or history. It is the part where theory is tested by prediction. It is the relation of evidential support, and its essentials can be schematized by means of little more than logical analysis" (1–2). The project of *Pursuit of Truth* sounds less like the undertaking announced in "Epistemology Naturalized"[1] and more like philosophy as the positivists understood it.

Quine's well-known holism also appears in a subtly moderated form. The radical thesis that the unit of empirical significance is the whole of science gives way to the modest claim that a testable set of sentences needs "critical semantic mass," which means it has to be "pretty big" (17). Holism corrects "the naive conception of scientific sentences as endowed each with its own separable empirical content" (16). But Quine distances himself from those coherence theorists, some inspired by Quine himself, who hold that all inference is inference to some total view of the world.

---

[1] In *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp 69–90.

0022-362X/92/8901/48–51

On the issue of meaning, Quine makes no retreats: determinate meanings are a myth. For a holist, "[t]rue sentences, observational and theoretical, are the alpha and omega of the scientific enterprise. They are related by structure, and objects figure as mere nodes of the structure" (31). Since "what is empirically significant in an ontology is just its contribution of neutral nodes to the structure of the theory," systematic reinterpretation of a theory's ontology gives us two theories empirically indistinguishable as wholes. This thesis of ontological relativity, or the inscrutability of reference, concerns only mutually compensating interpretations of words within sentences. But the more dramatic thesis of the indeterminacy of translation "declares for divergences that remain unreconciled even at the level of the whole sentence, and are compensated for only by divergences in the translations of other whole sentences" (50). Since meaning is whatever is captured by translation, if translation is indeterminate, so is meaning. But since all that matters in translation is "fluency and effectiveness of dialogue" (59), determinate meanings are dispensable, even for linguists and lexicographers.

Not so for irreducibly intensional discourse. Our practice of attributing beliefs and other propositional attitudes to our fellow speakers "harbors indispensable lore about human activity and motivation" (71). And though we do well to keep such talk out of science, which can then enjoy "the crystalline purity of extensionality" (71), there is no principled barrier to "science on the intensional side too, insofar as there are observational checkpoints however tentative" (72). The main thing is to avoid mixing vocabularies. Quine commends Donald Davidson's "anomalous monism" or "token physicalism": we can aspire to a fully extensional global theory of the world, but we still need an alternative, intensional way at looking at some of the events it covers. We may end up with "multiple scientific theories, jointly true" (73). (Truth, incidentally, is still just disquotation.) So where Quine once wrote of the irreducibility of the intensional as showing "the baselessness of intensional idioms and the emptiness of a science of intention,"[2] his attitude today is notably forgiving.

But while many of Quine's thoughts became clearer to me, some did not. In particular, I find Quine's conception of a "naturalized" theory of knowledge—its task and its relation to traditional epistemology—very murky. Quine has hinted both that naturalized epistemology replaces attempts to refute the skeptic and that it dissolves skeptical doubts. But here interest in skepticism is conspicuous by its absence, so matters are not clarified. And what he does say is not helpful. He repudiates "the Cartesian dream of a foundation for scientific method, firmer than scientific method itself." Nevertheless, he remains "occupied with what has been central to traditional epistemology, namely the relation of science to its sensory data" (19). But the traditional notion of "sensory data" is precisely that of a foundation for empirical knowledge. So I do not see how one can reject the dream of a foundation for science, while retaining an interest in anything like the traditional problem of connecting our knowledge of the world with our sensory data.

---

[2] *Word and Object* (Cambridge: MIT, 1960), p. 221.

I suspect that Quine has no clear conception of naturalized epistemology and that the source of his confusion is his use of "stimulation," which is a peripheral matter only physiologically speaking. A stimulation, defined as the set of all a given subject's triggered exteroceptors (3), is just a neurological event. But he also treats stimulations as the "data" from which we "project our theory of the external world." He sees his interest in "neural input" as epistemological, precisely because it concerns "the flow of evidence from the triggering of the senses to the pronouncements of science" (41). It is not easy to see how evidence can flow from things that are not even beliefs. To bear on the traditional problem of justifying our beliefs about the external world, stimulations must carry a distinctive content, graspable prior to knowledge of the world. This takes us back to the quest for a foundation.

For Quine, evidence proper comes in with observation sentences, which for the most part already concern external objects. Again, observation sentences may be prompted by stimulations but they are not "projected" from them. In fact, there is not even an explanatory, never mind a justificatory, connection between stimulations, classified according to physical criteria, and assent to observation sentences. Although acquiring an observational vocabulary may depend on a phase of conditioning to stimulation, once observation sentences become embedded in a set of sentences of critical semantic mass, the only condition under which they test anything, they cease to be under such tight stimulus control, for what stimulations prompt assent to a given observation sentence becomes sensitive to collateral knowledge. Holism intervenes: mature observation sentences have no empirical content, hence no stimulations, which they can call their own. Indeed, to try to explain beliefs by appeal to stimulations would be to mix extensional and intentional talk in the very way anomalous monism warns us against. As Quine himself notes, epistemologically relevant similarity, with respect to stimulations, is perceptual not receptual: similarity of input is determined by similarity of response.

Why the confusion? Because Quine wants to think that, purged of its dogmas, empiricism remains science's philosophy of science. Stimulations as data are a faint, lingering echo of the myth of the given. Thus Quine tells us that "the most notable norm of naturalized epistemology actually coincides with that of traditional epistemology": *nihil in mente quod non prius in sensu* (19). But by Quine's own showing, nothing epistemologically relevant is "in the senses," as Quine understands the senses. So this "watchword of empiricism" no longer carries its traditional sense, reducing to a banal warning against "telepaths and soothsayers." Only Quine's tendency to treat stimulations sometimes as mere neurological events and sometimes as constituting a kind of primitive, experiential knowledge, allows him to see a close connection between epistemology as he understands it and its traditional empiricist forbear. Resolve the tension and the connection disappears. Much of Quine's naturalized epistemology disappears with it, however.

Quine's holism also works against the idea that there is much to be said about scientific method. Logical analysis shows, as Karl Popper has insisted, that "[p]ure observation lends only negative evidence, by refuting an observation categorical that a proposed theory implies" (13). But holism ensures that there is never just one way of accommodating a recalcitrant observation. There are vague maxims, such as that of minimum mutilation. But there are no rules. Still less is there a science of how we project theories from data.

This is a short book. But I fear that for a philosopher who renounces empiricism's dogmas and shows no sustained interest in the roots of skepticism, epistemology is a short subject.

MICHAEL WILLIAMS

Northwestern University

Russell, Idealism, and the Emergence of Analytic Philosophy. PETER HYLTON. New York: Oxford University Press, 1990. xv, 420 p. Cloth $69.00.

At my university (Cambridge) the content of courses in the history of philosophy stops, conveniently, around the year 1900. Philosophers whose work comes after that date are dealt with in courses

0022-362X/92/8901/51–5

on "Metaphysics," "Logic," etc. As a result, our students read Bertrand Russell in the same context as that in which they read Donald Davidson; they engage with Russell as if he were an emeritus professor, only recently retired. This tradition owes much to Russell himself. It seemed to him that, as the new century began, he was escaping from the old cave of idealist appearances into the clear light of logic, mathematics, and, thanks to G. E. Moore, goodness. So although Russell thought that the works of previous philosophers were worth studying, the tradition that derives from his work—the analytic tradition in philosophy—cannot readily accept that by now his works just belong to the history of philosophy: for that might suggest both that his break with the past was not as sharp as it seemed, and that our current debates do not just continue his concerns. It might even suggest that, if he is the paradigm analytic philosopher, we are not.

Peter Hylton, himself a Cambridge graduate, aims precisely to insert Russell into the history of philosophy by acquainting us with his earliest works and refusing to engage in debates with him (except in the odd footnote). Thus the style of his book is altogether different from that of, say, Mark Sainsbury's *Russell*.[1] Whereas Sainsbury pursues themes from Russell down to contemporary debates in order to evaluate Russell's ideas from a current point of view, Hylton aims to set up a historical distance between Russell and ourselves so that we can appreciate the differences between Russell's views and our own. Admittedly, Hylton also wants to make contemporary analytic philosophers aware of those features of their methods which are historically explicable in the light of Russell's work. But his proposals in this area are rather sparse and questionable—e.g., that analytic philosophy has been decisively influenced by the thesis that "mathematical logic provides the crucial method of philosophy."

In his opening chapters, Hylton sets out to estrange us from the all-too-familiar author of "On Denoting" by introducing us to the unfamiliar territory of British idealism, as represented by T. H. Green, F. H. Bradley, and J. M. E. McTaggart. His aim is to get behind the superficial understanding of these philosophers which Russell's criticisms of them manifest, so that we can appreciate both Russell's own early writings, especially *The Foundations of Geometry*, and the complexity of the issues at stake in the subsequent debates between Russell and Bradley. Hylton indeed urges that many of Russell's criticisms of Bradley fail to engage with Bradley's thought because Russell already makes assumptions that Bradley would reject. Thus their debates about the reality of relations are largely pointless because Russell does not do justice to Bradley's conception of reality. By and large I agree with Hylton on this matter, though I do not think that even he succeeds in making the intellectual grounds for Bradley's ascent to the Absolute comprehensible. But there is one point which Hylton stresses which seems to me to need qualification. This concerns the issue of degrees of truth: Hylton takes this to be one of the fundamental disagreements between Russell and Bradley which is never properly thematized in their disputes. That there is a disagreement here is, of course, indisputable; but I think one's attitude to it is modified when one realizes that Russell's philosophy incorporates a variant of Bradley's relative/absolute distinction in its different attitudes to common-sense *analysanda* and to their philosophical analyses. Even a Moorean insistence on the certain truth of our common-sense judgments cannot disguise the fact that, once it is allowed that these judgments stand in need of analysis, their truth becomes relative to that of their analysis.

---

[1] New York: Routledge, 1979.

In describing Russell's rejection of idealism, Hylton rightly stresses the decisive role of the young Moore. Hylton characterizes Moore's early philosophy as "Platonic Atomism." The atomism is certainly there, and is central to Moore's rejection of Bradley's totalitarian holism. I am less happy about the Platonist comparison; for Moore did not think that there is a real world of forms (or concepts) behind ordinary appearances. Instead, he thought that reality extends beyond that which actually exists—the possible is as real as, and of the same ontological type as, the actual (so the appropriate comparison is with David Lewis). Hylton's misjudgment here leads him to overlook the basis of Moore's views about necessity, about which he puzzles a good deal. In fact, Moore just has a Leibnizian view, which he expresses in terms of the universality of necessary truths.

Hylton devotes his longest chapter to *The Principles of Mathematics*. He shows first how Russell combined Moore's philosophy with the new logical techniques he had learnt from Peano to put together a powerful anti-idealist polemic on the topic of infinity and thus "to play the part of Samson" in relation to Kant and his disciples. And on this matter Hylton does seem content to allow the validity of Russell's criticisms of idealism. But, Hylton argues, this judgment should be qualified by a recognition of the fact that Russell lacked a workable account of his own logical theory. The central issue here is that of "the variable," which Russell discusses in conjunction with his theory of denoting concepts and propositional functions. Hylton suggests that Russell had hoped to handle variables by simply treating the expression 'any term' as expressing a denoting concept, but found that his theory of denoting concepts was not adequate to the complexities of multiple quantification: as a result, Russell was compelled to treat the conceptions of a propositional function and of a variable object as fundamental, although he was not at all content with this conclusion. This seems to me largely correct: and Russell's discussion of these issues does now seem strikingly archaic, though I suspect that the techniques developed by W. V. Quine in "Variables Explained Away" (in his *Selected Logical Papers*[2]) could be exploited to give Russell's theory of denoting concepts a better chance of success.

The stage has now been set for "On Denoting." Hylton shows persuasively that the old theory of denoting concepts could handle the puzzles concerning empty names and referential opacity with which Russell introduces the new theory. So the motive for the new theory must come from some other dissatisfaction with the old theory. In part, it just must have been the perceived need anyway for a separate treatment of variables; but Hylton also provides a plausible reconstruction of the notorious "The First Line of Gray's Elegy" argument which both delivers the requisite conclusion, that the relation between a denoting concept and its denotation seems inherently inexpressible, and avoids the imputation of gross use/mention confusions. Although Hylton appears to think that this argument is satisfactory from Russell's point of view in 1905, I did not find it convincing, since the reconstructed argument just assumes that denoting concepts cannot ever occur as nondenoting terms (Hylton even cites a passage from an unpublished manuscript in which Russell appears to question this assumption). My own view (to which Hylton also alludes briefly) is that Russell was always under pressure to abandon the theory of denoting concepts because it was inconsistent with a fundamental premise of his "Platonic Atomism" —that true propositions do not represent, or denote, the facts— they are themselves the facts. Therefore, once Russell had worked out how to handle the problematic case of definite descriptions without the old theory, he was keen to abandon it altogether— though Hylton intriguingly suggests that the relation between a propositional function and its values is not so very different from that between a denoting concept and its denotation.

---

[2] New York: Random, 1966.

Russell's theory of descriptions leads into his conception of incomplete symbols and the theory of logical fictions. But this line of thought, central to Russell's analytic program, intersects with another that also has its origins in *The Principles of Mathematics*—the theory of types. Russell introduces this to cope with his paradox; in the context of *The Principles of Mathematics,* however, as Hylton shows, the theory conflicts with Russell's assumptions that anything can be a term, and that all terms have the same status. Since these assumptions are central to "Platonic Atomism," Russell's subsequent development of the theory of types marks a significant retreat from that philosophy, and in the last part of his book Hylton discusses similar points, such as the replacement of propositions by the multiple-relation theory of judgment and the abandonment of the naive direct realism characteristic of the epistemology of his early thought. Hence, as Hylton progressively reintroduces us to that old friend, the author of *The Problems of Philosophy,* we find we have to bid farewell to the stranger whose acquaintance we have just been making—the author of *The Principles of Mathematics.*

There is much in Hylton's long book that I have not discussed (e.g., his treatment of *Principia Mathematica*). I have concentrated on his central theme, the rise and fall of "Platonic Atomism," and in his discussion of this, Hylton, I think, succeeds brilliantly. As I have indicated, I have some reservations about details, but the overall conception and achievement is extraordinary. Hylton's book provides an exemplary combination of scholarship, lucidity, and imagination. He sets a new standard for all subsequent discussions of Russell's philosophy.

THOMAS BALDWIN

Cambridge University/Clare College

MEMO TO AUTHORS

The editors cannot afford to send to the printer manuscripts that are difficult for the compositor to read or footnotes that need to be completed in proof. They ask that manuscripts submitted be prepared according to the following guidelines:

1. All material, including the quotations and footnotes—especially footnotes—must be typed, double-spaced, on one side of the paper, with ample margins left and right, top and bottom.
2. No bibliographies will be published; bibliographical material should be put into the footnotes.
3. Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include city of publication, publisher, and date. Footnotes should be gathered at the end.
4. Quotations should be checked by the author; capitalization, italics, spelling, etc., should match the original. Corrections should not have to be made (expensively) in proof.
5. If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked; if there are many, a list should be provided. Standard logic symbols are available, and letters of the Greek and script alphabets; no boldface.
6. Computer copy can be extremely difficult to read; only letter-quality copy is acceptable. It should be double-spaced and not justified. The print should be dark enough and the letters widely enough spaced so that both letters and punctuation marks can be read easily.
7. Finally, photo copies on shiny paper are not acceptable. Ribbon copy is much preferred, but clear copies are acceptable.

The editors are willing to read and evaluate manuscripts that do not meet these standards, provided they are genuinely legible. They require retyping of accepted manuscripts, if necessary, before scheduling for publication.

The *Journal* is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally consider to no more than 7,500 words, or eighteen *Journal* pages. A more precise way of determining length is by characters: 2,850 characters per *Journal* page, or 51,300 characters per *Journal* article of 18 pages. [Calculate character count by counting the number of characters per line (including blank spaces and punctuation), multiplying by the number of lines on an average page, multiplying by the number of pages (footnotes included), and dividing by 2,850.] Manuscripts very much in excess of these limits will be returned unread.

The *Journal* does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere.

CALIFORNIA PHILOSOPHY

Risk and Rationality
Philosophical Foundations for Populist Reforms
K.S. SHRADER-FRECHETTE
Shrader-Frechette shows why extremist views about risk are unreliable and sketches the changes needed to set hazard assessment and risk evaluation on a publicly defensible foundation
$15 95 paper, $39 95 cloth, 272 pages

Essays on the Foundations of Aristotelian Political Science
Edited by CARNES LORD and DAVID O'CONNOR
These essays cover concepts such as production, race, class, and gender, as well as more traditional Aristotelian topics such as justice, monarchy and democracy *$39 95 cloth, 380 pages*

Theories of Justice
BRIAN BARRY
Now in paper—"In 400 closely reasoned and often luminous pages [Barry] goes a long way towards explaining both the impact of Rawls's book and some of the most puzzling features of its reception "— *Times Literary Supplement*
California Series on Choice & Political Economy
$14.95 paper, 443 pages

At bookstores or order toll-free 1-800-822-6657. Visa/MasterCard.

Just Doctoring
Medical Ethics in the Liberal State
TROYEN A. BRENNAN, M.D., J.D.
"A marvelously engaging book."
—Theodore R. Marmor, Yale University
This lively and controversial plea for a rethinking of medical ethics goes to the heart of society at the end of the twentieth century. *A Centennial Book*
$34 95 cloth, 301 pages

Max Weber and Thomas Mann
Calling and the Shaping of the Self
HARVEY GOLDMAN
Now in paper—"A first-rate work in German intellectual history."
—*American Historical Review*
$12.95 paper, 295 pages

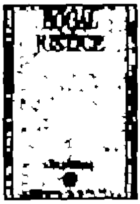
Fleeing the Iron Cage
Culture, Politics, and Modernity in the Thought of Max Weber
LAWRENCE A. SCAFF
Now in paper— "The best book about Weber yet written in English."
— *American Political Science Review*
$12 95 paper, 279 pages

Empirical Knowledge
ALAN H. GOLDMAN
Now in paper—"An enormously ambitious book It will make rewarding reading for anyone interested in these issues "—*Review of Metaphysics*
$14 95 paper, 420 pages

UNIVERSITY OF
CALIFORNIA PRESS
BERKELEY LOS ANGELES NEW YORK OXFORD

Oxford

Equal Justice



ERIC RAKOWSKI, *University of California, Berkeley*

The core of this book is a novel theory of distributive justice premised on the fundamental moral equality of persons. In the light of this theory, Rakowski considers three types of problems which urgently require solutions—the distribution of resources, property rights, and the saving of life—and provides challenging and unconventional answers.

1991 400 pp. $49.95

Confrontations with the Reaper

A Philosophical Study of the Nature and Value of Death

FRED FELDMAN, *University of Massachusetts, Amherst*

Presenting a clear and engaging discussion of the classic philosophical questions surrounding death, this book studies the great metaphysical and moral problems of death. Feldman concludes with a novel consequentialist theory about the morality of killing, applying it to such thorny practical issues as abortion, suicide, and euthanasia.

1991 264 pp. $24.95

Philosophical Interpretations

ROBERT J. FOGELIN, *Dartmouth College*

Placing particular emphasis on understanding the argumentative or dialectical role that passages play in their specific context, Robert Fogelin here collects fifteen of his essays, organized around the theme of interpreting philosophical texts.

1991 272 pp. $29.95

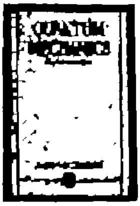
Law and Objectivity

KENT GREENAWALT, *Columbia University Law School*

In modern times the idea of the objectivity of law has been undermined by skepticism about legal institutions, disbelief in ideals of unbiased evaluation, and a conviction that language is indeterminate. Greenawalt here considers the validity of such skepticism and in the process, illuminates the development of jurisprudence in the English-speaking world over the last fifty years, assessing the contributions of many important movements.

1992 304 pp. $34.95

Quantum Mechanics



An Empiricist View

BAS C. VAN FRAASSEN, *Princeton University*

After introducing the empiricist point of view in philosophy of science, and the concepts and methods of the semantic approach to scientific theories, van Fraassen discusses three stages of quantum theory. After offering a critique of earlier interpretations, he develops a new one—the modal interpretation—which attempts to stay close to the original Copenhagen ideas without implying a radical incompleteness in quantum theory.

1991 560 pp.
paper $29.95 cloth $69.00

Locke and French Materialism

JOHN W. YOLTON, *Rutgers University*

This book tells for the first time the long and complex story of the involvement of Locke's suggestion that God could add to matter the power of thought in his *Essay Concerning Human Understanding*.

1991 256 pp. $55.00

✦*New in paperback!*

Marx and Ethics

PHILIP J. KAIN, *University of Santa Clara, California*

"Easy to read, interesting, and well-organized introduction to the issues surrounding the question of ethics and Marx....An important contribution to Marxian analysis."—*Contemporary Sociology*

1989 (paper 1991) 232 pp.
paper $19.95 cloth $49.95

Plural and Conflicting Values

MICHAEL STOCKER, *Syracuse University*

Plural and conflicting values are often held to be conceptually problematic, threatening the very possibility of ethics, or at least rational ethics. Rejecting this view, Stocker demonstrates why it is so important to understand the issues raised by plural and conflicting values and shows that plurality and conflict are commonplace and generally unproblematic features of our everyday choice and action.

1990 376 pp. $45.00

Morality and Moral Theory

A Reappraisal and Reaffirmation

ROBERT B. LOUDEN, *University of Southern Maine*

In this important new book, Louden responds to the arguments of both "anti-morality" and "anti-theory" skeptics, reaffirms the necessity and importance of moral theory in human life, and shows that moral theories fulfill a variety of genuine and indispensable human needs.

1992 272 pp.
paper $14.95 cloth $35.00

Love and Power

The Role of Religion and Morality in American Politics

MICHAEL J. PERRY, *Northwestern University School of Law*

"Perry's elegantly and eloquently written book should be essential reading for all who believe in human rights. He makes a compelling argument that a vibrant political discourse about the existence and nature of universal human rights must include religious beliefs about human nature and the human good."—Nadine Strossen, *American Civil Liberties Union*

1991 240 pp. $29.95

✦*New in paperback!*

The Limits of Morality

SHELLY KAGAN, *University of Illinois, Chicago*

"Kagan writes with exemplary clarity and shows considerable ingenuity in arguing for his massively counter-intuitive and really rather sinister thesis."—*Times Higher Education Supplement*

1989 (paper 1991) 432 pp.
paper $24.95 cloth $55.00

Conditionals

Edited by FRANK JACKSON, *Monash University, Melbourne*

Introducing some of the most interesting current work on conditionals, this book pays particular attention to possible world semantics for conditionals, the role of probability in helping us to understand conditionals, implicature and the material conditional, and subjunctive versus indicative conditionals.

1991 248 pp.
paper $14.95 cloth $39.95

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MG
To order by phone using major credit cards please call (212)679-7300, ext.7106

Oxford University Press
200 Madison Avenue ✦ New York, NY 10016

Oxford

On Rhetoric
A Theory of Civic Discourse
ARISTOTLE
Newly translated with Introduction, Notes, and Appendixes by
GEORGE A. KENNEDY

"Easy to read, trustworthy, and has the added distinction of bringing out more clearly than any other translation the way Aristotle's reasoning works. It will improve the general state of Aristotle studies in the English-speaking world."—Thomas Conley, *University of Illinois*
1991 (paper January 1992) 368 pp.
paper $9.95 cloth $27.95

Essays on the Philosophy of Socrates
Edited by HUGH H. BENSON

"An excellent collection...useful resource for students and scholars."
—Richard Kraut, *University of Illinois, Chicago*
1991 336 pp. paper $14.95
cloth $45.00

The Philosophy of the Visual Arts
PHILIP A. ALPERSON

"A scholarly, well-organized anthology—in fact the only one to provide a systematic introduction to the philosophical foundations of the visual arts—that will be equally useful to students and to scholars."
—Lydia Goehr, *Wesleyan University*
January 1992 528 pp.; 108 illus.
paper $29.95

Philosophical Issues in Journalism
ELLIOT D. COHEN

"Provocative readings, which should be noticed and used by journalism educators, on major conceptual issues of journalistic theory and practice."
—Deni Elliott, *Dartmouth College*
January 1992 272 pp. paper $14.95

Philosophy, Politics, Autonomy
Essays in Political Philosophy
CORNELIUS CASTORIADIS
Edited by DAVID AMES CURTIS

These remarkable essays by Cornelius Castoriadis include his latest contributions to philosophy, political and social theory, classical studies, development theory, cultural criticism, science, and ecology.
1991 320 pp. paper $14.95
cloth $35.00

New in paper
The Ethics of International Business
THOMAS DONALDSON

"The best thinking about the role of ethics in business....Integrating the teaching and practice of management with ethics and the humanities."
—*Ethics Digest*
1989 (paper January 1992) 222 pp.
paper $16.95 cloth $26.00

Reason and Religious Belief

An Introduction to the Philosophy of Religion

MICHAEL PETERSON, WILLIAM HASKER, BRUCE REICHENBACH, and DAVID BASINGER

"Makes some of the best recent work in philosophy of religion easily accessible to undergraduates."
—Philip L. Quinn, *University of Notre Dame*
1990 304 pp. paper $17.95

What is Justice?

Classic and Contemporary Readings

Edited by ROBERT C. SOLOMON and MARK C. MURPHY

"A superb collection of major treatments of justice, covering not only western texts but critical works from Islam and Chinese philosophy. Its contemporary selections are most representative of current scholarship. An excellent text for colleges."
—Stephen Pope, *Boston College*
1990 394 pp paper $15.95
cloth $39.95

The Nature of Mind

Edited by DAVID M. ROSENTHAL

"A unique and valuable collection of classical and contemporary works in the philosophy of mind."
—Robert Van Gulick, *Syracuse University*
1991 656 pp. paper $19.95
cloth $49.95

Pragmatics

A Reader

Edited by STEVEN DAVIS

"An extremely well-conceived collection of work at the cutting edge." —Harold Morick, *State University of New York, Albany*
1991 608 pp. paper $32.50

The Philosophy of Language

Second Edition

Edited by A.P. MARTINICH

"The best collection I know."
—Leora Weitzman, *University of Wisconsin, Madison*
1990 544pp. cloth $31.00

What Does It All Mean?

A Very Short Introduction to Philosophy

THOMAS NAGEL

"An excellent introduction."
—Robert M. Fowler, *University of Colorado*
1987 106 pp. paper $8.95
cloth $16.95

Seven Theories of Human Nature

Second Edition

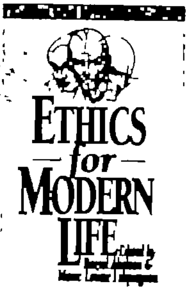
LESLIE STEVENSON

"Well organized, useful, systematic presentation of recent types of social theory. A fine starting point for consideration of current trends."
—Paul Mattson, *University of California, Santa Barbara*
1987 (paper 1988) 160 pp.
paper $9.95 cloth $24.95

Prices and publication dates are subject to change
To request an examination copy, write on school letterhead giving full course information, including course name, level expected enrollment, and your decision deadline to

Oxford University Press
ATTN: College Sales Coordinator
200 Madison Avenue • New York, NY 10016

ST. MARTIN'S PRESS



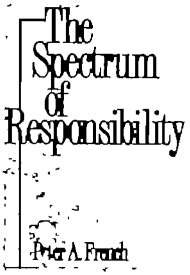
ETHICS FOR MODERN LIFE
Fourth Edition
RAZIEL ABELSON and MARIE-LOUISE FRIQUEGNON

A well balanced introduction to both theoretical and applied ethics that first examines the standards of moral right and wrong and then applies these basic philosophic principles to modern social issues. The chapter on environmental ethics is new to this edition, and the chapter on suicide has been reinstated from the second edition.
Paperbound. 480 pp. 1991

REALITY, KNOWLEDGE, & THE GOOD LIFE
A Historical Introduction to Philosophy
Willem A. deVries

REALITY, KNOWLEDGE, AND THE GOOD LIFE
A Historical Introduction to Philosophy
WILLEM A. DEVRIES

A text-reader containing 31 selections from 18 major philosophers. The text presents three major themes in the history of philosophy – reality, knowledge, and the good life – and shows how the themes have evolved over time in the writings of important philosophers from the early Greeks to the 20th century.
Paperbound. 704 pp. 1991



THE SPECTRUM OF RESPONSIBILITY
PETER A. FRENCH

A brief anthology that brings together differing viewpoints on the concept of moral responsibility – specifically, the responsibility of individuals and groups to each other and to society as a whole. Responsibility is examined at both the collective and personal level, offering students the opportunity to see moral roles in the business or professional world.
Paperbound. 236 pp. 1991

SMP
ST. MARTIN'S PRESS
College Division, Department JR
175 Fifth Avenue, New York, N.Y. 10010
College Desk 1-800-446-8923

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 2
FEBRUARY 1992

page

57 *Sour Grapes and Character Planning* Luc Bovens
79 *Loopholes in Moralities* Thomas W. Pogge

BOOK REVIEWS
99 Henry E. Allison: *Kant's Theory of Freedom* Paul Guyer

CALCUTTA UNIVERSITY CENTRAL LIBRARY

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 2, FEBRUARY 1992

SOUR GRAPES AND CHARACTER PLANNING*

> "Fit yourself into accord with the things in which your position has been cast, and love the men among whom your lot has fallen, but love them truely."
>
> Marcus Aurelius

As a freshman in college I was first exposed to the Stoic *Lebensregel* that one should mold one's desires such that they come to match whatever it is that life has in store. Aside from being distrustful, I remember being puzzled by this ancient advice. What is puzzling about it is that, though the advice is well-taken in certain cases, it is strictly off the mark in others. Imagine a person who switched her preference for pepsi or coke back and forth depending on what the nearest vending machine had to offer. Similarly, Nietzsche's Zarathustra "laughed at the weaklings who thought themselves good because they had no claws."[1] Yet, at the same time, there is something very respectable about the capacity to exercise control over one's own emotions and desires. It was argued by Harry Frankfurt[2] that freedom of the will precisely consists in being able to control what desires will issue in action. And what

---

* This paper is a shortened version of the first chapter of my Ph.D. dissertation, *Reasons for Preferences*, University of Minnesota/Minneapolis. I also thank the National Fund for Scientific Research (Belgium) and the National Science Foundation (U.S.A.) for their financial support. I am grateful to the numerous individuals who have made contributions to this paper: C. A. Anderson, George Bealer, Willem deVries, Kevin Falvey, Ronald Giere, Jean Ladrière, Andrew Mason, H.E. Mason, Michael Otsuka, Michael Root, Naomi Scheman, and Yutaka Yamamoto.

[1] *Thus Spoke Zarathustra*, in *The Portable Nietzsche*, W. Kaufmann, ed. (New York: Viking, 1954), p. 230.

[2] "Freedom of the Will and the Concept of a Person," this JOURNAL, LXVIII, 1 (January 14, 1971): 5–20.

0022-362X/92/8902/57–78

could be wrong with choosing one's desires so as to avoid disappointment and frustration?

This tension in the Stoic advice is captured very crisply in Jon Elster's[3] discussion of sour grapes and character planning. These phenomena both involve a shift of desires in response to what one takes to be feasible or morally correct. For less respectable cases of this kind, the term *sour grapes* (SG) comes to mind. For more respectable cases, Elster reserves the term *character planning* (CP). He goes on to ask what distinguishes SG from CP. It is this question which articulates my earliest apprehension over the Stoic advice. Elster appeals to the autonomous nature of CP that is lacking in SG. I shall question this stand and defend an alternative solution that runs parallel to Donald Davidson's account of *akrasia*.

I

Elster extensively discusses the question what preferences must be like in order for them to be rational. Preferences acquired through SG are a paradigm case of irrational preferences. The fox in La Fontaine's fable has an appetite for grapes, though when he finds out that he cannot reach for them, he claims that grapes are too sour for his canine taste anyway. (On this reading of the fable, the fox does not change his *beliefs* about the sourness of the particular grapes in question, but rather changes his *tastes* concerning the sourness of grapes in general.[4]) Elster contrasts preference acquisition through SG with preference acquisition through CP. It is a commonplace in Aristotelian ethics that an action cannot qualify as a virtuous action unless the agent performs it with a sense of joy. This sense of joy is acquired through habituation. Imagine a poker player who enjoys the game tremendously because of the opportunities for cheating involved. She comes to realize that cheating is morally reprehensible and decides to start a better life on this score. Initially she does not find fair-play poker terribly exciting. Nonetheless, in her quest for Aristotelian virtue, she is committed to becoming the kind of person who enjoys fair play. And in order to carry through this project of CP, she chooses the Aristotelian route of habituation. That is, she decides not to give in to her initial misgivings and to stick to fair play, hoping that some day she will thus come to enjoy this style of playing.

This Aristotelian scheme can be extended beyond the moral realm. The poker player's choices were constrained by moral considerations. But choices may also be constrained simply because alternatives are no longer feasible. Imagine that the management of the casino decided to install proper lighting, such that the opportunities for cheating are entirely blocked. Our poker player is not at all pleased with this new arrangement. Nonetheless, she decides to stick with it and to try to develop a liking for fair play. It may well turn out that, after a few games, she will indeed come to enjoy fair-play poker quite a bit better than her former style.

---

[3] *Sour Grapes* (New York: Cambridge, 1983), pp. 20–6, 117–9; *Explaining Technical Change* (New York: Cambridge, 1983), pp. 84–8; *Rational Choice*, Elster, ed. (New York: Blackwell, 1986), pp. 14–5.

[4] Cf. *Sour Grapes*, p. 123n

What is puzzling about SG and CP is that both phenomena can be described in the same way, namely, both the fox and the poker player adjust their preferences in reference to what they believe to be feasible alternatives. Nonetheless, while the fox is an epitome of irrational preference acquisition, the same judgment could hardly be passed on to the poker player. Nothing stands in the way to say that her preferences took shape in an entirely rational fashion. So what is it about SG and CP that makes us pass such radically different judgments on these phenomena? The phenomena fit the same description, so what is left out of this description which can account for this distinction? Before I turn to Elster's solution to this puzzle, I shall shortcut a few incorrect, though nonetheless very telling suggestions.

(a) It may be suggested that our puzzle is ill-phrased in that it assumes there actually are two phenomena in play. One might argue that 'SG' and 'CP' describe one and the same phenomenon, which is irrational in some respects and rational in other respects. If one wants to highlight the irrational character of a particular preference adjustment in reference to the feasible alternatives, then 'SG' is the appropriate term. If one wants to highlight the rational character of the very same case, then 'CP' is the term that comes to mind. I admit that, for a great many cases of preference adjustment in reference to the feasible alternatives, it is not clear whether to categorize them as instances of SG or as instances of CP. Arguments can be made on either side of the fence and no empirical information is lacking which could decide the debate. One cannot deny, however, that there are also clear-cut cases around. It is simply incorrect to describe the fox's preference adjustment as an instance of SG. The existence of such clear-cut cases is sufficient to safeguard our puzzle. This is nonetheless a very telling suggestion. A correct account of SG and CP should also explain why there is such a broad "gray" area between both phenomena, i.e., why it is that so many cases seem to have something of SG as well as something of CP about them.

(b) It may be suggested that, whereas the poker player takes her time to come to appreciate fair play, the fox renounces his appetite for grapes at the spur of the moment. But how could this difference in timing account for our judgments concerning the irrational character of SG and the rational character of CP? Furthermore, we can well imagine a fox who passes through a phase of denial, anger, or resignation before he finally makes the SG move or a CP poker player who is very adaptable and in no time develops an appreciation for fair play. Nonetheless, this suggestion is very telling in that CP typically involves more of a project than SG. If a preference for an alternative that is no longer feasible is abandoned at the spur of the moment, there is a presumption of irrationality. This is not the case if such involves a gradual process. A correct account of SG and CP should also provide an insight into what underlies this presumption.

(c) It may be suggested that, whereas the poker player actually did change her preference, the fox merely engages in an act of self-deception. The phenomenon of SG does not consist in a preference adjustment, but rather in constructing a false belief about a stable preference for a nonfeasible alternative. I admit that there are genuine cases of self-deception in which a person forms a false belief about her preference for a nonfeasible alternative. In such cases, one will come to abandon this false belief upon sincere introspection or would indeed choose for the alternative if it were suddenly made available. Yet I want to claim that there are cases in which a person actually does change her preference because some alternative is no longer feasible and which do not qualify as instances of CP. Suppose that, upon sincere introspection, one would still hold on to one's adjusted preference or, if the nonfeasible alternative suddenly became feasible once again, one would turn away from it. I want 'SG' to refer to such cases of irrational preference acquisition, not to cases in which a false belief is being formed about a stable preference. Upon sincere introspection, the fox would still claim to have lost his appetite for the grapes and, if we were to lower the vine for him, he would turn away his head. Nonetheless, there is something very telling about this suggestion, in that, for each case of SG, there is a persistent temptation to give a description in terms of self-deception about a fixed preference. A correct account of SG should be able to explain this temptation.

II

In trying to solve the puzzle, Elster argues that preferences are rational only if they are rationally acquired and the criterion for rational preference acquisition is autonomy. This criterion is meant to distinguish SG from CP. Unlike the fox's whims, the poker player's preference reversal in favor of fair play is autonomous in character. What is distinctive about autonomous preference acquisition is that it is open to intentional explanation.[5] Elster distinguishes between three types of scientific explanation, namely, causal, functional, and intentional. He appeals to Davidson's theory of action to fill in his account of intentional explanation.

Some of our ***doings***, i.e., some of the things we do, are ***actions***. Actions are doings that are intentional under some description. Other doings, like tripping over the carpet, are not intentional under any description. These are not actions, they are ***mere doings***. Davidson argues that

> (I) A doing (and hence an action) is intentional under some description *d*, if and only if
> - (a) there is some subset of the agent's mental states which provides for a reason for the doing under *d*,
> - (b) these mental states were causally efficacious in bringing about the doing,
> - (c) and this causal-efficacity is of the correct type.

The set of mental states, which provides for a reason for an action under some description, contains a cognitive component (a belief) and an evaluative component or a ***pro attitude***. For instance, a reason for the action under the description 'adding sage to the stew' is the agent's pro attitude or desire to improve the taste of the stew together with the belief that adding sage to the stew will improve its taste.[6]

Let us return to Elster's theory. A doing is open to intentional explanation if and only if there exists a description under which the doing is intentional. Preference acquisition is a doing for both the fox and the poker player. Here is how the distinction comes in between SG and CP. The poker player's preference acquisition is open to intentional explanation, i.e., there exists a description of her preference acquisition under which it is intentional. She intentionally adjusts her preferences in reference to the feasible alternatives. The fox's preference acquisition is closed to intentional explanation. He does not intentionally adjust his preferences in reference to the feasible alternatives, nor does there exist any other description under which his doing is intentional. Consequently, the poker player's preference acquisition is an action, while the fox's preference acquisition is a mere doing. We can thus generalize: preference acquisition through CP is an action, while preference acquisition through SG is a mere doing.

---

[5] ***Explaining Technical Change***, pp. 84–8.

[6] *Actions and Events* (New York: Oxford, 1980), pp. 4–12, 78–9, 86–7. Notice that Davidson himself does not use the expressions 'a doing' and 'a mere doing'.

What accounts for this distinction? Elster argues that the former phenomenon is caused by a metadesire to adjust one's preferences, while the latter is caused by an "affective drive"[7] that aims at reducing the frustration from unsatisfiable preferences. Metadesires, unlike drives, are pro attitudes and thus may qualify as an evaluative component in the set of mental states which provides a reason for an action under some description. Drives, on the other hand, can be no more than causes of mere doings. But what is so different about drives and desires which could motivate this distinction?

Elster believes that drives differ from desires in that they "are not conscious and known to the person who has them."[8] I have first-person knowledge of my desires as the reasons for my actions. I do not have first-person knowledge of my drives as the causes of my doings. Secondly, unlike drives, desires "may forgo short-term pleasure to achieve some longer-term gain."[9] Prudence can be an attribute only of desires, not of drives. And thirdly, Elster more than once takes recourse to mechanistic metaphors to refer to drives. According to Davidson, common mental events are in principle open to both a full physical description and a full mental description (*op. cit.*, pp. 245–9). Elster believes that drives are open only to a full physical description. Neuroscience does not yet allow for such description, however, and hence for the individuation of drives we must take recourse to their behavioral effects. If we want to describe the causal agents of these behavioral effects as such we can only invoke metaphors, like "forces" or "the wirings of the pleasure machine."[10]

Elster admits that his characterization of the set of autonomous preferences is merely tentative. First, he wants to exclude preferences shaped by metadesires that are themselves nonautonomous in character. Second, he acknowledges that some preferences are fully autonomous, though they were not intentionally acquired. He mentions socialization within a particular culture and learning to appreciate alternatives by simply trying them out as nonintentional processes of preference acquisition. One could hardly deny that at least some preferences so acquired are autonomous in character.[11]

---

[7] *Sour Grapes*, p. 24.

[8] *Ibid.* Cf. *ibid.*, p. 21; *Rational Choice*, p. 15; *Explaining Technical Change*, p. 87

[9] *Sour Grapes*, p. 25.

[10] *Sour Grapes*, p. 25; *Explaining Technical Change*, pp. 71–2, with reference to A. Tversky, "Self-deception and Self-perception: Some Psychological Observations," paper presented to a colloquium on "The Multiple Self" (Paris: Maison des Sciences de l'Homme, 1982).

III

The core of Elster's theory is that preferences are rational only if they are rationally acquired and that preferences are rationally acquired if and only if they are autonomously—that is, intentionally—acquired. Since in CP preferences are intentionally acquired and in SG they are not, the former preferences are rational, the latter are not. I shall first argue that Elster's understanding of intentional preference acquisition is not in line with his professed commitment to Davidson's analysis of intentional action, nor with his own commitment to rational-choice theory. Subsequently, I shall challenge intentional preference acquisition on a strict Davidsonean reading as a criterion for rational preferences.

Consider the following application of Davidson's analysis (I) to intentional preference acquisition:

> (I^{pa}) A person intentionally acquires a preference if and only if some subset of her mental states provides for a reason for her acquiring the preference and is causally efficacious to this effect in the correct way.

Elster argues that preference acquisition through drives is not intentional, since drives are not conscious and known to the person who has them. This suggests a commitment to the claim that a person intentionally acquires a preference only if she has first-person knowledge of her reasons for doing so. Yet this claim is not in line with Davidson's analysis of intentional action. Rather, it suggests the following requirement:

> (i) An action is intentional under some description *d* only if the agent has first-person knowledge of her reasons for doing *d*.

Davidson argues convincingly against the inclusion of this condition in his analysis of intentional action.[12] Secondly, claim (i) is incompatible with two other claims to which I believe Elster is committed:

---

[11] *Sour Grapes*, pp. 22, 112–4, 138–9; *Explaining Technical Change*, pp. 84–7.

[12] Davidson argues that a person may intentionally poison Charles, though nonetheless be mistaken about her reasons for doing so. For instance, she may believe that her reason for doing so is to save Charles pain, while her actual reason is to have him out of the way (*op. cit.*, pp. 17–8).

(ii) Rational-choice theory can tell us more about why some person performed an action than a genuinely honest actor could (I take 'rational-choice theory' to stand for the body of theories that explain actions by means of decision-theoretic or game-theoretic models);

(iii) Explanations in rational-choice theory are paradigmatic cases of intentional explanation.

Elster is clearly committed to (iii).[15] Consider claim (ii). Many successful explanations by means of decision-theoretic or game-theoretic models unveil preferences of which the agents do not care to have first-person knowledge. The agents tell more flowery stories about the springs of their action and, what is more important, they are genuinely honest when they tell these stories. Considering the actual practice of rational-choice theory, I believe that Elster will acknowledge that this theory brings out springs of actions of which the agent does not have first-person knowledge.

Now consider these three claims. If rational-choice theory provides for a form of intentional explanation (on (iii)), it lays out the springs of actions that are intentional under some description. On (ii), these springs are at least sometimes beyond one's first-person knowledge. On (i), they must be within one's first-person knowledge. This is how the inconsistency comes in. Since Elster is committed to (ii) and (iii), he cannot hold that first-person knowledge of the reasons for one's actions is a necessary condition for intentional action. Hence, he cannot include first-person knowledge of the reasons for a preference acquisition as a necessary condition for intentional preference acquisition.

Let us now respect Elster's professed commitment to Davidson's analysis of intentional action and consider whether analysis (I^{pa}) could be a plausible criterion in Elster's theory.

First there is a challenge of principle. Preference acquisition is a type of doing. We ascribe rationality or irrationality not only to preference acquisitions, but also to doings in general. There are standard patterns in our practice of making such ascriptions to doings in general, e.g., if a doing is an action and is caused by an irrational belief (say, a belief acquired through wishful thinking), then we call the doing itself irrational. Since preference acquisitions are types of doings, a criterion for the ascription of (ir)rationality to preference acquisitions may be more specific than a criterion for the ascription of (ir)rationality to doings in general. The former more specific criterion cannot violate the latter more general criterion, however. A criterion for the ascription of (ir)rationality to preference acquisitions is constrained by standard patterns in our common practice of ascribing (ir)rationality to doings in general.

---

[15] There is Elster's explicit avowal that rational-choice explanations are a variety of intentional explanations (*Rational Choice*, p. 12). Furthermore, rational-choice theory is the focal point of Elster's discussion of intentional explanation (*Explaining Technical Change*, pp. 74–83).

Let 'a description *i*' stand for a description of an action under which the action is intentional. Consider the following claims concerning the ascription of (ir)rationality to doings in general.

(i) Consider the set of all descriptions *i*. The descriptions in this set all denote actions that are intentional under *i*. Within this set, there is ***both*** a subset of *i*s which denote actions that are rational under *i* and a subset of *i*s which denote actions that are irrational under *i*.

(ii) Consider the set of all descriptions of doings. This set contains:
  (S_1) the subset of descriptions denoting actions that are intentional under this description;
  (S_2) the subset of descriptions denoting actions that are not intentional under this description;
  (S_3) and the subset of descriptions denoting mere doings.
  Outside (S_1) there simply is no description of a doing that is either rational or irrational under that description.

Elster clearly underwrites claim (i). Certainly an action may be rational under some description *i*. Furthermore, much of Elster's work is devoted to developing a taxonomy for actions that are irrational under some description *i*. For example, if the set of mental states which provides for a reason for an action under some description *i* contains a belief that is irrational because it is an instance of wishful thinking, then the action itself is irrational under this description.[14]

Let us turn to claim (ii). First, consider descriptions of mere doings, i.e., doings that are not actions, like tripping over the carpet. Certainly it is meaningless to say of such doings under any description that they are either rational or irrational. Secondly, consider descriptions of actions that are not themselves descriptions *i*. In 1978, Albino Luciani (John-Paul I) died less than 5 weeks after his election as pope of the Roman Catholic Church. The sentence 'The conclave elected Albino Luciani' and 'The conclave elected a pope who had less than 5 weeks to live' describe the same action, though only the former sentence is a description *i*. It may well have been rational or irrational for the conclave to elect Albino Luciani, though it would be meaningless to ask whether it was rational or irrational for the conclave to elect a pope who had less than five weeks to live.

---

[14] *Sour Grapes*, pp. 26, 123–4; *Rational Choice*, pp. 13–4.

On Elster's criterion, preferences are acquired rationally if and only if they are acquired intentionally. This criterion violates the constraints on the ascription of (ir)rationality in claim (i) and (ii). On (i), if preferences are acquired intentionally, then such action may be rational or irrational under this description. On claim (ii), if preferences are acquired nonintentionally, then it is meaningless to say that such doing—regardless of whether it is an action or a mere doing—is rational or irrational under this description.

Two other challenges are pragmatical in character. Elster spells out a necessary condition on rational preferences on the basis of what he takes to be characteristic differences between SG and CP. Two questions can be asked. First, does this necessary condition—namely, intentional acquisition—indeed point to a characteristic difference between preferences acquired through SG and preferences acquired through CP? Secondly, we also hold judgments concerning the rationality of preferences outside the context of SG or CP. Does intentional acquisition indeed express a necessary condition on rational preferences in general?

Considering the former challenge, I fully agree that preference acquisition through CP is intentional in character. But why would preference acquisition through SG be any less so? Elster argues that the latter phenomenon is caused by a drive and that a drive is not a mental state and hence cannot qualify as a pro attitude in the reason for an action under some intentional description. Yet why is it that drives do not qualify as mental states? Elster mentions the absence of first-person knowledge for drives. On my earlier argument, this is a moot point, of course, since it may well be the case that a person does not have first-person knowledge of the mental states that provide for a reason for an action under some intentional description. Elster also mentions the imprudent character and the orientation toward short-term pleasure of drives.[15] Yet mental states may well share these features. My passion to cruise the Caribbean is not any less of a mental state, furnishing the pro attitude in the reason for my intentional purchase of a ticket to Barbados, when I do *not* have the necessary funds as when I *do* have the necessary funds. Hence, I cannot see any good reason why preference acquisition through SG is any less intentional in character than preference acquisition through CP.

---

[15] Notice that this characterization is inconsistent with Elster's inclusion of the phenomenon of "forbidden fruit is sweet" as a type of preference acquisition through drives (*Sour Grapes*, pp. 111–2). Clearly such counteradaptive preferences are not geared toward (short-term) pleasure.

Let us now turn to the latter pragmatical challenge. Must rational preferences in general be intentionally acquired? Elster admits that there are counterexamples to his theory. There exist preferences which are indeed nonintentionally acquired, but which are autonomous and hence fully rational. He mentions preference acquisition through socialization or through learning. I take these counterexamples to be more than an indication that the theory is in need of some fine tuning.

I have restricted CP to the project of adjusting one's preferences to the feasible alternatives through the medium of habituation. Notice, however, that there are many other media for intentional preference acquisition. One such medium is precisely learning. Here is an example. I do not like dry wine, though I want to develop an appreciation for it, since all my wine connoisseurs friends serve only dry wines at their dinner parties. In order to do so, I may decide to enroll in a wine-tasting course, rather than fill up my cellar with dry wines. That is, I may choose to mold my preferences through the route of learning, rather than through the route of habituation. There is nothing irrational about preferences that are intentionally acquired through the medium of learning. But neither is there anything irrational about preferences that are nonintentionally acquired through learning. I may well come to like dry wines in a wine-tasting course, without ever having planned such preference change. This suggests that intentional acquisition may have very little to do with the ascription of rationality to preferences.

The same kind of argument can be made for other media of preference acquisition as well. CP is intentional preference acquisition through habituation. But, similarly, there is nothing irrational about nonintentional preference acquisition through habituation. After some years in North America, peanut butter has slightly moved up from the bottom of the list in my evaluation of alternative sandwich spreads. I have never planned such a thing, yet there is nothing irrational about this nonintentional preference change through habituation. In conclusion, intentional preference acquisition does not seem a promising candidate as a necessary condition on rational preferences in general.

IV

In this section, I shall venture a new solution to our puzzle. I share Elster's intuition that the phenomenon of SG yields irrational preferences while the phenomenon of CP typically[16] yields rational preferences. The challenge thus consists in constructing a necessary condition on rational preferences, such that CP can pass the test, while SG does not.

Preferences are defined over alternatives, but they are not the only kind of evaluative judgment defined over alternatives. Consider the following instructive quote from Davidson:

> . . . particular choices can be explained in decision theory along lines very similar to those followed in reason explanations: a particular action is chosen from among the available set because of the agent's beliefs (for example how apt he thinks the action is to produce various results) and the relative values he sets on the possible outcomes. His desires are thus made comparative and quantitative. Some ordinary desires, however, do not translate directly into preferences, so that not all reason explanations have a clear decision theory counterpart explanation. This is particularly evident if we think of conflicting desires. A person may have a reason for preferring *A* to *B* and another reason for preferring *B* to *A*. This is an embarassment for reason explanations, for they need to predict which reason will win out. Decision theory skips the problem, for it says nothing about why one basic outcome is preferred to another, and the theory bars possible evidence of conflict in behaviour (*op. cit.*, p. 269).

Ordinary expressions of wants, i.e., of evaluative judgments defined over alternatives, are ambiguous. If a person claims to want some alternative at some point in time, it is unclear whether she is acknowledging that she has reasons that favor the alternative—reasons that may or may not be trumped by competing reasons at that point in time—or whether she is indicating that the alternative ranks highest in her all-out ranking at that point in time. For instance, suppose I am on a strict diet and utter the words 'I really want a chocolate malt'. This may be taken to mean that I am strongly attracted to the sweet flavor of chocolate malt but may nevertheless stick to my commitment to dieting. Or it may mean that I have decided to slip on my diet and that chocolate malt has moved to the top in my all-out ranking over alternative desserts. Instances of the latter type of evaluative judgments are named "preferences" in decision theory and are represented formally by utility functions. Let us name instances of the former type of evaluative judgments *de-*

---

[16] I introduce this qualification because we want to allow that some cases of CP derive their irrationality from familiar patterns of irrationality, e.g., when a character planner would act on irrationally acquired beliefs.

sires. Hence, desires allow for evaluative conflict, while preferences do not. In other words, I may desire alternative *A* over alternative *B* (for some reason) *and* I may desire *B* over *A* (for some other reason), but I cannot have a preference for *A* over *B* *and* have a preference for *B* over *A*.[17] So what is the status of preferences within the enclosing set of evaluative judgments?

In "How is Weakness of the Will Possible?" (*op. cit.*, pp. 37–42; also 97–102) Davidson sketches an analysis of practical reasoning. In this analysis he constructs a taxonomy of evaluative judgments defined over actions. This taxonomy will be helpful in trying to understand the place of preferences within the set of evaluative judgments defined over alternatives.

Evaluative judgments defined over action tokens are either relational or nonrelational judgments. Relational judgments defined over action tokens are relative to reasons. Let *a* and *b* be two action tokens. For some reasons, I may judge that it is better to do *a* than *b*, while, for some other reasons, I may judge that it is better to do *b* than *a*. On a logical analysis, the main connective in a relational judgment is the two-place sentential operator '. . . is a ground to judge that ———', in which the first place is filled in by a set of reasons and the second place by a ranking of action tokens.

Nonrelational judgments defined over action tokens are not relative to reasons. They are the intentions that accompany our actions. If I intentionally do *a* rather than *b*, then my action is accompanied by the intention to do *a* rather than *b*. Davidson takes this intention to be the nonrelational judgment that *a* is better than *b*.

Reasons for relational judgments defined over action tokens come in pairs, each containing a relational judgment defined over action types and a belief ascribing action tokens to action types. Here is an example. On this cold winter night, I am considering whether I shall walk my dog or read a book in front of the fireplace. On the one hand, I may think it is better to walk my dog than read a book, considering my dog needs the exercise. On the other hand, I may think it is better to read a book, considering it is so cold out. On a logical analysis, my reason to think it is better to walk my dog than read a book is the relational judgment defined over action types:

---

[17] Davidson's claim that a person may have a reason for *preferring* *A* to *B* and another reason for *preferring* *B* to *A* is slightly confusing. While *preferences* in a technical sense do not allow for evaluative conflict of this kind, Davidson apparently does not carry over this technical sense from 'preference' to the related verbal form 'preferring'. I assume that 'preferring *A* to *B*' is here being used in a nontechnical sense and is not the kind of evaluative judgment that is at work in a decision theory, i.e., it is not synonymous with 'having a preference for *A* over *B*' in a decision-theoretic sense.

(M_1) That some action token x is an instance of the action type 'giving proper care to pets' and some action token y is an instance of the action type 'neglecting pets' is a ground to judge that x is better than y.

and the belief:

(m_1) The action token 'my walking my dog (now)' is an instance of the action type 'taking proper care of my dog' and the action token 'my reading a book (now)' is an instance of the action type 'neglecting pets'.

The conclusion of this reason pair is the relational judgment defined over action tokens:

(C_1) (M_1) and (m_1) are a ground to judge that the action token 'my walking my dog (now)' is better than the action token 'my reading a book (now)'.

Analogously, a conflicting relational judgment defined over action tokens (C_2) can be formed which is relative to the relational judgment defined over the action types 'exposing myself to the cold weather' and 'not exposing myself to the cold weather' (M_2) and the matching belief ascribing the action tokens in question to these action types (m_2):

(C_2) (M_2) and (m_2) are a ground to judge that the action token 'my reading a book (now)' is better than the action token 'my walking my dog (now)'.

There is a particular relational judgment defined over action tokens in which we take a special interest, namely, the relational judgment that is relative to *all* of the agent's reasons, i.e., all of the mental states that the agent takes to be relevant to the choice at hand. Such relational judgment is an *all-things-considered judgment*. Singular reason pairs may lead to conflicting relational judgments defined over action tokens. The all-things-considered judgment is then determined by the relative importance that the agent assigns to each reason pair. All-things-considered judgments do not follow from the agent's reasons as a matter of logic. I myself am a pet lover and the cold weather does not throw me off that easily. So here is my brave all-things-considered judgment:

(C_{all}) M_1, m_1, M_2, and m_2 are a ground to judge that 'my walking my dog (now)' is better than 'my reading a book (now)'.

Nonetheless, I may stay home and read a book in front of the fireplace. How so? Either I may not have formed an all-things-considered judgment at all, because I did not give sufficient thought to the choice at hand. Or I may have formed such judgment but, through weakness of the will, I did not act in accordance with it. In both cases, the intention or nonrelational judgment that accompanies my action is not informed by the all-things-considered judgment, but rather by the relational judgment (C_2), which is relative to a subset (containing M_2 and m_2) of the complete set of reasons.

On Davidson's principle of continence, an action is rational only if the intention that accompanies it is informed by the agent's all-things-considered judgment for the choice at hand. Both ill-considered as well as *akratic* actions fail to pass this criterion of rational action.

What can be learned from Davidson's taxonomy for understanding the place of preferences in the set of evaluative judgments defined over alternatives? Let alternatives be *state-of-affairs tokens* (SOA tokens). For instance, my preference for this particular Volvo over that particular Austin *is* my preference for the SOA token of 'my owning this Volvo (now)' over the SOA token of 'my owning that Austin (now)'.

Preferences are a special kind of judgment defined over SOA tokens in that they do not allow for evaluative conflict. For some reasons I may desire the Volvo rather than the Austin and for some other reasons I may desire the Austin rather than the Volvo, but I cannot have a preference for the Volvo over the Austin and have a preference for the Austin over the Volvo. This should ring a bell. I propose that preferences, just like intentions, are nonrelational judgments. Other evaluative judgments defined over SOA tokens are relational judgments, i.e., they are relative to reasons. And this parallel can be carried even further. Reasons for relational judgments defined over SOA tokens come in pairs, each containing a relational judgment defined over SOA types and a matching belief ascribing SOA tokens to SOA types. For example, a reason for desiring this Volvo rather than that Austin is the relational judgment defined over SOA types:

($M^{\dagger}_{1}$) That a SOA token x is an instance of the SOA type 'owning a Swedish car' and a SOA token y is an instance of the SOA type 'owning a British car' is a ground to judge that x is better than y.

and the belief:

> (m_1^*) The SOA token 'my owning this Volvo (now)' is an instance of the SOA type 'owning a Swedish car' and the SOA token 'my owning this Austin (now)' is an instance of the SOA type 'owning a British car'.

This reason pair implies the relational judgment defined over SOA tokens:

> (C_1^*) (M_1) and (m_1) are a ground to judge that the SOA token 'my owning this Volvo (now)' is better than the SOA token 'my owning this Austin (now)'.

Again, different reason pairs may support conflicting relational judgments defined over SOA tokens. A particular relational judgment in which we take a special interest is the all-things-considered judgment defined over SOA tokens. This judgment is relative to *all* of the person's reasons, i.e., all of the mental states that the person takes to be relevant to the ranking at hand.

Davidson's principle of continence imposes a constraint on actions. This principle can also be understood to impose a constraint on the intentions accompanying the actions. The intention to do *a* rather than *b*, i.e., the nonrelational judgment defined over action tokens that it is better to do *a* than *b*, is rational only if it is informed by the all-things-considered judgment for the choice at hand. I propose to construct an analogous constraint for preferences:

> (R) A preference for SOA token *s* over SOA token *t*, i.e., the nonrelational judgment defined over SOA tokens that *s* is better than *t*, is rational only if it is informed by the all-things-considered judgment for the ranking at hand.

All-things-considered judgments are relational judgments that are relative to all of the person's reasons, i.e., all of the pairs—each containing a relational judgment defined over SOA types and a belief ascribing SOA tokens to SOA types—which the person takes to be relevant to the ranking at hand. Let us name a relational judgment defined over SOA tokens which is relative to *one* such reason pair *a criterial judgment*. An example of such criterial judgment is (C_1^*), i.e., on a more colloquial version, my judgment that, considering its country of origin, I would rather have the Volvo than the Austin. But there may also be other criteria, which the person takes to be relevant to her ranking, say, color, year, etc. The role of the all-things-considered judgment is to arbitrate between criterial judgments. If all criterial judgments favor the same alternative, arbitration simply means acknowledging the unanimous vote. If different criterial judgments favor different alternatives, arbitration involves an assessment of the relative importance that each consideration carries in one's mental life. I can thus rephrase the constraint on preferences:

> (R′) A preference for SOA token *s* over SOA token *t* is rational only if it results from proper arbitration over ***all*** criterial judgments defined over *s* and *t*, implied by the reason pairs that the person takes to be relevant to the ranking at hand.[18]

How do we go about judging the propriety of such arbitration? For our purposes, it is sufficient to see that there are clear cases of preferences that are irrational precisely because they did not result from proper arbitration over all criterial judgments. For instance, suppose I get carried away by the wonderful stereo equipment in the Austin. I thus set my preference for the Austin over the Volvo, overlooking that the Austin is an older, British car in an ugly ochre. On due reflection, I must admit that the consideration that set my preference carries much less importance for me than the considerations that were overridden by it. It can be agreed that such would indeed be an irrational preference.

I shall now argue that this necessary condition on rational preferences can solve our puzzle, i.e., can drive a wedge between SG and CP. Let alternatives *s* and *t* both be feasible alternatives to begin with. I have a preference for *s* over *t* and this preference is informed by an all-things-considered judgment, i.e., it results from proper arbitration over all criterial judgments. It now turns out that, for some reason, alternative *s* is no longer feasible anymore. In response, I intentionally adjust my preference such that I come to have a preference for *t* over *s*. So far, SG and CP run entirely parallel. They also run parallel in that I may or may not have first-person knowledge of my reasons for adjusting this preference, i.e., I may or may not know that I resort to this preference adjustment *because* I realize that my favorite alternative is no longer feasible anymore.

---

[18] One may object that no more is required for rational preferences than that one properly arbitrate over some subset of the relevant criterial judgments which is commensurate with the prominence that holding preferences over the alternatives in question has in one's life. This objection is well-taken and can be accommodated for in criteria (R) and (R′) and in the subsequent discussion at a serious cost of stylistic simplicity, though without affecting the general argument.

The phenomena do differ, however, in the following respect. In a typical case of SG, I adjust my preference, though I do *not* adjust any of the relational judgments defined over the SOA types, nor the relative importance I attach to each consideration. Assuming fixed beliefs, the original set of reason pairs thus remains constant and there is no change in the relative importance that each pair carries in my mental life. Since the all-things-considered judgment for the ranking at hand is relative to this complete set of reasons, it too remains fixed over the course of the preference adjustment. In other words, nothing changes about the arbitration problem over the criterial judgments for the ranking at hand. Hence, the adjusted preference cannot be informed by the all-things-considered judgment, i.e., it cannot result from proper arbitration over all criterial judgments. And this is what accounts for the irrational character of preferences acquired through SG.

A typical case of CP is the more involved project in which I adjust my preference as well as my reasons for the ranking at hand. That is, either I shift around some of the relevant relational judgments defined over SOA types (which, together with fixed beliefs, imply a change in criterial judgments), or I change the relative importance that I attach to each consideration. Consequently, relative to my adjusted set of reason pairs and the relative importance I assign to them, I now think that the still feasible alternative *t* is better than the no longer feasible alternative *s*. In my adjusted all-things-considered judgment, alternative *t* moves up above alternative *s*. Hence, over the course of the CP move, my preference in flux remains informed by the all-things-considered judgment in flux. In other words, my preference results at all times from proper arbitration over the relevant criterial judgments. And this is why there is nothing irrational about preferences acquired through CP.

Let us turn to an example. For quite a while I have been interested in a particular new job, which would give me a chance to work overtime at a decent pay and with challenging career opportunities. I then find out that this change of jobs is no longer feasible for some reason. In response, I shift around my preference such that I come to prefer my actual job to the new job.

In a typical case of SG, this preference adjustment is the only evaluative shift in my mental life. I still hold the relational judgment over SOA types that 'having a job with opportunities for well-paid overtime work' is better than 'having a job which leaves time for leisure' and that 'having a job with challenging career opportunities' is better than 'having a more stable and less stressful job'. The set of relational judgments defined over SOA types and of beliefs ascribing SOA tokens to SOA types ('my having the new job' is an instance of 'having a job with opportunities for well-paid overtime work', etc.) still imply a pair of criterial judgments, which cast a unanimous vote in favor of the new job. I still am the kind of person who has all reason to prefer the new job to my actual job. If I do nonetheless have a preference for my actual job over the new job, then my preference is not informed by the all-things-considered judgment for the ranking at hand. In other words, my preference does not result from proper arbitration over the relevant criterial judgments, which, in this particular case, would be acknowledging their unanimous vote. Consequently, my adjusted preference in favor of my actual job can save me from frustration only at the cost of being irrational.

In a typical case of CP, I effect a more radical change in my mental life. Not only do I come to hold a preference for my actual job, but I also bring about a shift in the set of relational judgments defined over SOA types such that, indeed, I become the kind of person who genuinely has reason to hold the adjusted preference. I now appreciate the charm of a job which allows me more time for leisure and in which there is no pressure for promotion. My adjusted preference is thus informed by the all-things-considered judgment that is relative to my newly acquired relational judgments defined over SOA types together with fixed beliefs. That is, it results from proper arbitration over the newly acquired criterial judgments, which now cast a unanimous vote for my actual job. Hence, my adjusted preference in favor of my actual job passes through as a fully rational preference.

What about La Fontaine's fox? The fox loses his appetite for grapes when he finds out that he cannot reach for them. We do not assume, however, that he has changed his relational judgment defined over SOA types that, say, 'indulging in juicy summer fruits' is better than 'keeping to a more wholesome diet'. Over the course of the SG move, he remained the kind of fox who has a craving for juicy summer fruits in general. And what about the poker player? The CP poker player did not only change her preference concerning fair-play poker in the well-lit Casino, but she also came to hold the relational judgment defined over SOA types that 'playing a fair game' is better, is more fun than 'cheating'. Over the course of the

CP move, she has become the kind of person who genuinely values fair play in general.[19]

Both the unsuccessful job hunter and the poker player changed around *all* relevant relational judgments defined over SOA types in their CP moves. Notice that CP may well take a more subtle course. It may be sufficient to change a proper subset of the relevant relational judgment defined over SOA types to effect the desired shift in the ranking of the all-things-considered judgment. Or, in case of conflicting criterial judgments, it may be sufficient to adjust the relative importance attached to the respective considerations to have proper arbitration result in the desired preference shift.

V

In section I, I considered three responses to our puzzle. I argued that, though none of these responses was correct, they were all in some way very telling suggestions. A proper solution to our puzzle should provide some insight as to why it is that each suggestion does indeed seem prima facie plausible. Let us consider whether my theory of rational preferences can meet this challenge.

(a) Between clear-cut cases of SG and CP, there is a broad "gray" area of preference adjustments that are undecided as to whether SG or CP is the more appropriate classification. Can my theory account for this phenomenon? In a typical case of SG, my complete set of reasons for the ranking at hand remains untouched over the course of the preference adjustment. In a typical case of CP, I adjust both this set of reasons and my preference such that the adjusted preference is informed by the all-things-considered judgment that is relative to the adjusted set of reasons. In between these typical cases, there is a continuum of preference adjustments, in which I actually do shift around *some* relational judgments defined over SOA types or make *some* changes in the relative importance I attach to each consideration, though none of this is quite enough to let the balance in the all-things-considered judgment tip over in favor of the feasible alternative. Some steps are being taken in adjusting the set of reasons for the ranking at hand, though not quite enough for the adjusted preference to pass through as a rational preference. Between typical cases of SG and typical cases of CP there is a continuum of preference adjustments accompanied by not entirely successful shifts in the set of reasons for the ranking at hand.

---

[19] There are certain interesting worries involving (a) an attempt to blur the distinction between CP and SG by allowing for relational judgments that favor feasible SOA types, (b) radical evaluative shifts that are nonetheless irrational in character, and (c) shifts from irrational to rational preferences which are motivated by feasibility considerations. I intend to address these objections in a later paper.

Are preference adjustments on this continuum watered-down cases of SG or of CP? This is merely a verbal point. On the one hand, it seems appropriate to talk about SG for all cases on the continuum, because of the irrational character of the thus acquired preferences. On the other hand, if a person has made serious changes in her relational judgments defined over SOA types, or in the relative importance she attaches to the criteria involved, it seems appropriate to talk about CP, even though these changes may not be quite sufficient to support her preference adjustment. Many real-life or fictional cases of preference adjustment are actually located on this continuum. They are neither typical cases of SG, nor typical cases of CP. My theory shows how there is something to be said for classifying such cases in either category. And this accounts for the phenomenon that there is indeed a broad range of cases in which we are undecided as to whether 'SG' or 'CP' is the more appropriate term.

(b) The SG move *typically* is a response at the spur of the moment, while CP *typically* is more of a long-term process. If preferences are abandoned at the spur of the moment, there is a presumption of irrationality, which there is not if such occurs gradually. My solution to our puzzle can readily account for this presumption. While SG targets only one's preference, CP focuses on the complete chunk of one's mental life which is relevant to the preference in question. SG involves one single adjustment, CP involves a project of change. A sudden change of preference in response to considerations of feasibility thus makes us suspicious that SG rather than CP is in play.

(c) In trying to give an account of SG, there is always the temptation to call in self-deception, even so if the fox would deny his appetite for the grapes upon sincere introspection and would turn his head if we were to lower the vine for him. Can my solution account for this temptation?

In attributing a preference to a person we rely on presumptions. One such presumption is that, if a person acknowledges a preference upon sincere introspection, then she does actually hold such preference. Another presumption is that a person's choices are indicative of her actual preferences. Certainly, both presumptions may clash and, within particular circumstances, one presumption may seem to deserve more credibility than the other. For instance, if the controversial choice is made under stress, then we tend to put more trust in the introspective presumption. On the other hand, if acknowledging some preference is threatening to one's self-image, then we tend to put more trust in the behavioral presumption.

There is also a third presumption that comes into play, namely, a principle of charity in practical reasoning[20]: if a person holds a set of reasons which makes it rational for her to hold one preference rather than some other, then there is a presumption that she actually does hold this preference. In claiming that the fox actually did change his preference, this presumption of rationality is not respected. On the other hand, if we were to claim that the fox did not change his preference, then we would not respect both the introspective and the behavioral presumption.

I do not want to suggest that the number of presumptions can tip the scale. Rather, I propose to restrict the term 'SG' to those circumstances in which we tend to put our trust into the introspective and the behavioral presumption, rather than into the presumption of rationality. But even though one may feel most at ease in trusting the introspective and the behavioral presumption, the temptation still lingers to interpret the phenomenon in reference to a principle of charity in practical reasoning, i.e., to deny that any preference adjustment has taken place. If this temptation were to be followed, one would need to take recourse to self-deception to make up for not respecting the introspective presumption. Hence, my theory of rational preferences drives a wedge between SG and CP without describing the former phenomenon in terms of self-deception. At the same time, it can explain why there is a prima facie plausibility to such description.

LUC BOVENS

University of Colorado/Boulder

---

[20] Davidson, *op. cit.*, pp. 221–2; and "Paradoxes of Irrationality," in R. Wollheim and J. Hopkins, eds., *Philosophical Essays on Freud* (New York: Cambridge, 1982), pp. 289–305.

LOOPHOLES IN MORALITIES*

MORALITIES, in one sense of this word, are abstract standpoints outside the world from which what occurs within it—actions, persons, institutions, states of affairs—can be evaluated. In another sense, moralities are themselves in the world: more or less unified sets of moral beliefs and attitudes held by particular persons or groups. In this second sense, moralities have real effects; and these may themselves be made a subject of evaluation. Such a perspective upon conduct-guiding structures of values and norms—or *codes*—is familiar to economists and legislators, who take social effects into account in designing or evaluating a tax code, for example. My aim is to extend this perspective to both key domains of morality: to *ethics* (concerned with the moral evaluation of conduct and character, with the life one should lead and the person one should strive to be) and to *justice* (concerned with the moral evaluation of social institutions that regulate and pattern human interactions). My general question is then: What role, if any, should an evaluation of the effects of moral codes play in our substantive moral reflection?[1]

Clearly, we should take account of such effects only if their connection to the relevant code is reasonably firm. It is not rational to worry about any fortuitous effects a morality (in the second sense) may have had in the past, if they tell us nothing about what effects it is likely to have in the future. Taking account of the differential effects of alternative moralities makes sense only insofar as we can estimate these effects in advance. My question then concerns effects

---

* This essay has greatly benefitted from criticisms and suggestions by many persons, including Bruce Ackerman, Rogers Albritton, Christian Barry, Rüdiger Bittner, Alex George, Robert Guay, Amy Gutmann, Charles Larmore, David Luban, Sidney Morgenbesser, Warren Quinn, and Ling Tong.

[1] The notion of a code is closely linked to the idea of codification, and it may therefore seem surprising that I speak of moral codes. After all, it is one distinctive feature of the moralities with which we are familiar in the modern world that, unlike the law, they are *not* codified, nor officially adjudicated and enforced. We live, one might say, in a world without moral codes. But then my talk of moral codes is a response to a theoretical need rather than to an empirical reality. We cannot engage in any meaningful analysis of the effects of moralities unless we can fix upon some rather determinate specimen to which this analysis can be, somewhat rigorously, applied. Talk of moral codes then reflects an idealization. I shall ask you to consider hypothetical cases in which determinate moral codes are counterproductive. We may conclude upon reflection that adherents of such a code have reason to revise it. And this conclusion may have implications for ourselves, if our moral beliefs—though amorphous and diverse—are sufficiently similar in content and effects to the hypothetical moral code of the example.

0022-362X/92/8902/79–98

of moral conceptions which are predictable at least in a rough probabilistic way.

This may be reminiscent of Nietzsche's question: What is the value of our morality? Yet I am not judging a morality by some novel, alien measure of value, but by its own standards: Have we organized our moral commitments in a way that reflects, and helps effectively achieve, what by their own lights matters? My goal is to uncover new kinds of considerations of relevance to our moral reflection and discourse, which may help break the stalemate of conflicting appeals to moral intuitions and thereby make moral dialogue more fruitful, moral convergence more likely, and our moral values more effective.

To keep my topic manageable, I confine myself here to one class of predictable effects, namely, to cases where a code is *counterproductive* by giving *incentives* toward conduct that is *regrettable* by the code's own lights.[2] In saying that a moral code can be counterproductive, I am assuming that it is understood as having some point, rationale, purposes, or objectives. This is the typical case in the modern world. But it is also possible, of course, to think of a moral code as the moral truth or God's law. We can then think of it as not being in the business of producing anything, and hence as incapable of being counterproductive. On this understanding, we would not have second thoughts about our morality even if our holding it predictably led to the extinction of the human species or the disappearance of this morality itself.

That a code gives someone an *incentive* to act in a certain way means only that it gives her a reason to act in this way. Such a reason for conduct need not affect conduct itself: a code can encourage conduct without triggering or influencing conduct.[3] I count both the incentives a code gives and any conduct these may elicit as effects of this code. My general question has thus become more specific: Can the fact that a morality gives incentives toward regrettable conduct constitute at least some reason to revise this morality?

I believe so. But let me add three cautionary notes to this claim. First, the fact that a code gives an incentive toward conduct that in some instances is regrettable in some respect does not automatically render this code counterproductive *tout court*. The conduct in question may be (by the lights of the same code) valuable in other instances or in other respects and thus not be regrettable on the whole. Alternatively, the overall point, rationale, purposes, or objectives of the code may demand that it give the relevant incentive, or that it have certain features of which this incentive is an unavoidable byproduct, despite the fact that the incentive is also regrettable, even on the whole. The problems I want to examine are raised only when a code is counterproductive *all things considered*, i.e., when its overall point, rationale, purposes, or objectives would be better fulfilled if this code were replaced by a somewhat different code. I therefore assume in my more abstract discussions, or seek to show in my more concrete examples, that there really are no redeeming virtues that would allow one to say that the moral code under discussion—though it gives regrettable incentives—is not counterproductive all things considered.

---

[2] I shall skip the phrase 'by the code's own lights' wherever possible. But it must be kept in mind that we are throughout dealing with counterproductivity, i.e., with cases where moralities undermine their own effectiveness or, as it were, regret their own effects.

[3] For example, a tax deduction for charitable gifts provides an incentive to many taxpayers. But for some of these, the incentive may not tip the scales—they are more inclined than they would otherwise be, but still not inclined enough, to give. And others may not know about the tax deduction or may not realize how they can take advantage of it. Codes may then give incentives that do not affect conduct at all. As I shall show, some such incentives are problematic nevertheless.

Second, the effects of a code that is prevalent in some society depend to some extent on the life context of this society (including its culture, technology, and natural environment). Likewise, the effects of a code held by an individual are shaped by the life context of that particular person.[4] It may then sometimes be possible to reduce or eliminate the regrettable effects of a moral code through deliberate modifications of the life context in which it is held. There are other cases, however, where such strategies are either unavailable or themselves very costly in moral terms.

Thus, my claim specifically targets cases where a morality is counterproductive all things considered in virtue of giving regrettable incentives that are not reasonably avoidable. In such cases we have reason to revise our morality. But—and this is the third cautionary note—such a reason may not be decisive. I do not claim, by any means, that the effects of moralities are all that counts in moral reflection. Only that they are important, and decisive in some cases.

I. A CRUCIAL DISTINCTION

It may seem that an incentive given by a code can be problematic only insofar as it actually elicits regrettable conduct. This view may indeed be plausible as regards *concrete incentives*, i.e., reasons for conduct which a code gives to fully specific agents who actually do or will exist (or to statistical aggregates of such agents, such as the French taxpayer). But one can speak of incentives in another sense as well. *Ideal incentives* are encouragements or discouragements that a code, as it were, "wants" or "intends" to give (which does not mean that they must have been thought of, let alone intended, by some creator(s) of it). Less metaphorically, ideal incentives are reasons for conduct which a code would provide to persons in the given life context, if they were *ideal adherents*, i.e., agents who are fully rational, who fully understand the code and interpret it correctly, who are fully committed to comply with any requirements of the code, and whose circumstances and remaining interests fall within the normal range. Insofar as a code provides ideal incentives, we can say that it *guides* persons toward certain conduct (or misguides them, if this conduct is regrettable on the whole).

---

[4] These two paradigm cases do not exhaust the realm of possibilities. A moral code may be socially prevalent in a small social system, may be held by a minority within a society (with the majority perhaps committed to some other code), may be held by an individual who is highly influential in a society (the President of the U.S.A.), and so forth. In all these cases, the effects of the code are shaped by the life context in which it is operative.

Ideal incentives are of two kinds. With *compliance incentives*, ideal adherents are motivated solely by their commitment to do what the code requires or suggests. Incentives deriving from the code's requirements are always decisive for an ideal adherent (assuming the code is consistent), while incentives deriving from its suggestions need not be. With *reward incentives*, ideal adherents are motivated by other, code-independent interests of theirs insofar as these can be pursued without violating requirements of the code. Here a code encourages conduct by affecting the official pay-offs: eligibility for benefits or liability to burdens. Reward incentives need not be decisive.

Two examples: a tax code requiring that one pay taxes on income from nonmunicipal bonds provides a compliance incentive to pay taxes on such income, and a reward incentive to invest in municipal rather than nonmunicipal bonds; a morality that requires single young adults to volunteer for a stint in the military provides a compliance incentive to enlist, and a reward incentive to get married early.[5]

Such ideal incentives may manifest themselves in concrete incentives (and thereby possibly in conduct as well) if ideal adherents of the code or persons sufficiently similar to them actually exist. But ideal incentives may also fail so to manifest themselves: neither the compliance incentives nor the reward incentives in our two examples would manifest themselves as concrete incentives to persons who do not know the relevant parts of the codes or do not care whether they are in compliance or not.[6]

---

[5] I am here assuming that ideal adherents, having normal interests, care more about enhancing their own income than the government's and would also rather avoid enlistment. In the second example, serving in the military is morally required and legally voluntary.

As a code may provide ideal incentives that do not manifest themselves as concrete incentives, it may also give concrete incentives that do not correspond to any ideal incentives it provides. Thus, a tax law may incline certain persons to keep their fortune abroad (where it can more easily be hidden from the tax authorities). And our ethic may incline certain persons to pretend to be needy (in order fraudulently to exploit the ethical command to help the needy). In these cases, a code gives persons reasons to transgress a requirement of this code, which would not be reasons for them if they were ideal adherents of it.

With this distinction between the ideal and the concrete sense, our more specific question splits into two. One is the ideal question: May we have reason to revise our morality if, in the given life context, it provides ideal incentives that are regrettable by its own lights? The other is the concrete question: May we have reason to revise our morality if, in the given life context, it provides concrete incentives that are regrettable by its own lights? This essay focuses exclusively on the former question.[7]

II LOOPHOLES

Even if they affect no one's conduct, regrettable ideal incentives can be problematic by revealing a structural flaw in the code providing them. Suppose some tax code provides a compliance or reward incentive that would induce some taxpayers, if they were ideal adherents, to act in ways that directly undermine a central purpose of this code. The code may then be criticized for containing a loophole and thus being in a sense incoherent. And this critique can be sustained even if corresponding concrete incentives do not exist or are not noticed, so that the loophole does not actually elicit any regrettable conduct at all.

A *loophole* exists only if the connection between the code and the regrettability of the conduct it encourages is tight in two respects. First, the relevant incentive must be an ideal one, so that the code can be said to *guide* agents toward the regrettable conduct. Second, the conduct it encourages must be regrettable *in itself* rather than in virtue of any further effects it may bring about, however predictably. So, through loopholes agents are guided toward conduct that is in itself regrettable on the whole.

---

[6] By providing ideal incentives, a code guides (or misguides) agents. Like a traffic sign, it can do so without affecting their conduct: it is possible that agents differ from ideal adherents, so that the code's ideal incentives do not manifest themselves as concrete incentives to them. And it is possible that agents, however similar to ideal adherents, disregard reward incentives or compliance incentives deriving from the code's suggestions.

[7] Some preliminary thoughts on the latter question are presented in my "The Effects of Prevalent Moral Conceptions," in *Social Research*, LVII, 3 (Fall 1990): 649–63.

Conduct is regrettable in itself just in case it goes against what the code requires, or essentially aims at outcomes that are (by the code's own lights) regrettable on the whole. By definition of "ideal adherent," a consistent code will not provide *ideal* incentives toward conduct that is in the first way regrettable in itself—it will not, that is, guide agents toward conduct that it also forbids. The topic of loopholes thus boils down to cases where a code guides agents toward conduct that essentially aims at outcomes that are regrettable on the whole. I call such ideal incentives *inherently regrettable.*

III. SOCIAL ARRANGEMENTS

To illustrate the problem, let me here concentrate on moral loopholes involving *social arrangements,* by which I mean relationships, associations, or social systems, such as marriages, lawyer-client relationships, families, firms, churches, and states. Social arrangements tie their members together through normative expectations that are based upon special ties, including moral ties, that define special rights and obligations, powers and responsibilities. Such special moral ties vary in formality—ranging from the justiciable and enforced explicit provisions of a legal code to the vague understandings that give rise to moral expectations (and disappointments) among friends or compatriots.

Consider the ethical view that as a member of a social arrangement one may sometimes (when acting in behalf of other members or of the entire group) deliberately harm outsiders in some specific way, even while one may not do so when one is acting on one's own. Such views are, I believe, actually held. In the business world, those who implement a corporate policy that is harmful to consumers, employees, or the general public often stress their status as managers and their obligations toward the firm's owners, whose financial interests they were hired to promote. How is this supposed to be ethically relevant? Such a manager is not trying to pass the buck by suggesting that the owners, too, must bear some of the responsibility for the offensive policy and that he therefore bears less. What he implies, rather, is this: thanks to the differentiation of roles between owners and manager, the policy in question can be implemented without wrongdoing—though this would not be the case if he were managing his own firm so as to promote his own financial interests.[8]

---

[8] Similarly, in the academic world, university presidents and trustees have sometimes defended their decision to invest part of their institution's endowment in

Ethical views of this sort guide their adherents to form or join social arrangements in order to effect, through the special ties these involve, a unilateral reduction of responsibility toward those left out of these arrangements. Such guidance is inherently regrettable—not, of course, because additional social arrangements may be formed, but because harms that are straightforwardly morally precluded can be indirectly brought about after all. Let us consider two examples in some detail.

IV. CASE 1

Your father owns a large apartment complex. Renting the apartments as they are is profitable, but market conditions are such that it will soon become preferable to convert them into luxury accommodations while doubling the rents. Your father knows, however, that most of the apartments are occupied by elderly tenants who have lived there for decades. Few, if any, of them could afford to stay. The others would have a hard time finding new homes. More importantly, they would be scattered all over town and thus would lose the companionship they have found with one another—the greatest source of pleasure in their final years. Your father understands that it would be wrong of him to force the tenants out. But he also resents the idea of giving up the extra income. It occurs to him that many landlords like himself employ lawyers to manage their real estate in the owners' best interest. So your father hires a lawyer as well and instructs her to manage his real estate in a professional way. As was to be expected, the lawyer eventually initiates the conversion of the complex. Rents double and the old tenants are scattered all over town.

The assumptions I am making about the prevalent ethical code, *E*, are indicated in the top row of the following table:

	Option A1	Option A2	Option B
(aggregate) moral status:	permitted	forbidden	permitted
aggregate nonmoral cost:	high	zero	negligible

E forbids your father to initiate the conversion on his own (A2). As for the remaining two options, *E* does not require (though it may suggest) abandonment of the conversion (A1) in preference to the lawyer's initiation of the conversion in the owner's behalf (B). A

---

morally questionable ways by appealing to their fiduciary obligations toward the university and its members. They imply that an investment (e.g., in South African goldmines or in certain arms-export firms) can be ethically acceptable when it is made on behalf of the university even while it would not be ethical if it were made with personal funds.

landlord is not required to be personally involved in the management of his property. He may surely entrust his real estate to a lawyer. And such a lawyer is not required to refrain from the conversion. She has a professional responsibility to her client, and therefore may (or even should) administer property entrusted to her in the client's best interest. It is simply not her role to ponder whether her client is not rich enough as it is, and should not therefore forego some benefit in order to preserve the tenants' community and form of life.[9]

An ethic incorporating these beliefs provides an ideal incentive to hire the lawyer.[10] This incentive is inherently regrettable, provided the responsibility of the "solitary" landlord is conceived as, to some extent at least, a responsibility *toward his tenants*, and the wrong he would commit by converting the building (A2) as a wrong done *to them*. I assume here that one main rationale for *E*'s forbidding such single-handed conversions is to protect the tenants' community against deliberate destruction for the sake of greater landlord profits. *E* itself suggests this rationale if, for example, its evaluation of the single-handed conversion varies with the extent to which it would foreseeably harm the tenants. Insofar as this rationale is present, the incentive in case 1 indicates a loophole: *E* guides the landlord toward the profitable conversion by enabling him to render it free of wrongdoing. Yet, according to *E*'s rationale, as implicitly suggested in *E*, this conversion is regrettable on the whole: the tenants' suffering is not outweighed by the landlord's extra income. *E*'s own rationale thus gives a reason for closing the loophole—for blocking the opportunity unilaterally to reduce aggregate responsibility.

This does not mean, of course, that an acceptable ethic must require a lawyer to give no more weight to her client's interests than to those of other persons. It is not objectionable that persons may form or join social arrangements whose members agree to owe one another much more than persons owe one another in general as a matter of common decency. So I have no quarrel with the idea that persons, by becoming members of a social arrangement, may *increase* what they owe the other members and may thus come to owe them more than they owe all outsiders. What I find problematic is the idea that persons, by so increasing what they owe certain others, may *reduce* what they minimally owe everyone else. I say minimally, because an ethic may of course allow persons to reduce what they owe others; it may, for instance, allow one to get a divorce through which one reduces what one owes to the person one had married. So my point concerns only those basics which persons owe all others as a matter of common decency. However an ethic may fix these minimal general demands, it should not enable persons unilaterally to reduce or dilute them. Specifically, it should not allow them, by forming or joining a social arrangement, to subject themselves to new, countervailing obligations to its other members which may outweigh, trump, or cancel their minimal obligations to everyone else. Forming or joining a social arrangement should not have that sort of moral efficacy, should not introduce that sort of moral change. It should not be among the legitimate benefits of forming or joining a social arrangement that one thereby weakens the minimal moral protections that those who are now outsiders enjoy against oneself.

---

[9] It is important that the lawyer, too, is in compliance with what *E* requires (that the *aggregate* moral status of option (B) is "permitted"). I am not interested here in the opportunity, which some ethical codes may afford, of reducing one's own wrongdoing (one's share of responsibility for some transaction) by partly shifting it off to another.

[10] More precisely, such an ethic—as sketched for example in Charles Fried, "The Lawyer as Friend: The Moral Foundations of the Lawyer-Client Relation," in *Yale Law Journal*, LXXXV (1976)—gives landlords an ideal compliance incentive to prefer (B) over (A2) and an ideal reward incentive to prefer (B) over (A1).

Even this narrow demand is defeasible because an incentive—even though it depends on the opportunity unilaterally to sidestep aggregate responsibility for given harms—can be desirable in another respect and so not regrettable on the whole. About case 1, for example, someone might object that I have underdescribed *E*'s rationale: there is a point after all to the incentive *E* provides, namely, the eminent moral importance of the hiring of lawyers. It matters more, morally, that landlords should hire lawyers than that they and their lawyers should not knowingly destroy their tenants' community for extra profit. This objection marks a weak spot in my entire approach, because, quite generally, the claim to have found a loophole can always be countered by claiming that the relevant ideal incentive is not in fact regrettable *on the whole*.

Still, I am not too worried about this weak spot, because what counts is not making a counterclaim but making it plausible. To see whether it is plausible, we must examine whether the code overall really reflects the suggested rationale for the questionable incentive. If it were part of *E*'s rationale that additional hiring of lawyers is morally important, then this value ought to surface throughout the code. If there are no other indications of such a value in the code, then the conjecture proposed by the objector is implausible.

Insofar as *E* is merely a creature of my own imagination, there is no point in conducting such an examination for case 1. But insofar as *E* is meant to be a realistic code that may actually be held in our society, we can dismiss the counterclaim. For no one (except a few lawyers?) seriously believes it to be morally important that there be more hiring of lawyers.

I conclude that, if *E* prohibits self-initiated conversions at least partly on account of the damage they would do to the tenants' form of life, for the sake of greater landlord profits, then any ideal incentive *E* provides toward such hiring of lawyers indicates a loophole. Such an incentive manifests a kind of incoherence: the point of forbidding a given deliberately imposed correlation of harms and benefits in one context is defeated, for no compelling reason, when this prohibition can be avoided by switching to another context in which the same deliberate imposition of the same correlation of harms and benefits is permitted.[11] To avoid this incoherence, *E* must not allow aggregate responsibility for the conversion, or aggregate wrongdoing involved in initiating the conversion, to be so significantly reduced by the hiring of a lawyer.[12]

V. CASE 2

The year is 2100. Like their ancestors, South Africa's whites are quite affluent thanks to an abundance of black workers available at subsistence wages, who now, however, are nationals of sovereign states. This change was instituted by an earlier generation of South African whites, who intended that their progeny should not only (like themselves) be economically advantaged but should also be free from any special responsibility for the fate of blacks in the region. The former white leaders (in the 1980s) keenly appreciated that the prevailing conception of justice, *J*, allows even its minimal constraints on economic relations to be undermined through the interposition of national borders. And so they made sure the blacks of the twenty-first century would be foreigners.

The incentive operative here is again a ***moral*** incentive, in that the creation of the homelands is motivated by a feature of *J:* by the weight *J* gives, in its assessment of economic relations, to the existence of states and of ties among compatriots. It corresponds to an

---

[11] The phrase 'for no compelling reason' alludes to the possibility of plausible counterclaims. If additional hiring of lawyers were morally important, and if the incentive of case 1 were a good way of promoting this important value, then there might be a compelling reason for the incentive.

[12] You may feel that it was pretty clear, even before my argument, that a plausible ethic should not condone your father's conduct. But we shall soon see (sections VI and VII) that the problem of loopholes requires a far deeper revision of *E*. Moreover, I have specifically chosen *E* for the purpose of giving a plausible illustration of a constraint that I believe should be imposed upon *any* moral conception As will emerge, this constraint poses a radical challenge to other moral views that are commonly held.

ideal (compliance) incentive, because the reform instituted by the former white leaders really relieves the injustice (in terms of *J*) in the relations between whites and blacks in southern Africa, and thus would matter even to persons who are genuinely committed to *J* and aware of all relevant facts. Finally, insofar as *J* condemns the earlier pattern (prevailing in the 1970s) on account of the disadvantaged position of blacks, the incentive it provides is also ***inherently regrettable:*** through the encouraged "reform," those disadvantages get preserved, even entrenched, rather than eradicated.

A conception of justice providing this incentive would, I think, be incoherent in a sense. *J* strongly disapproves of the pattern that existed in the old South Africa (of the 1970s) as violating basic justice: blacks are, by law or convention, barred from the more lucrative positions in their economy; they compete for the remaining positions at a great disadvantage due to grossly inferior schooling, discriminatory hiring and promotion practices, legal prohibitions against residing in certain areas, and so on. *J* then approves of the "reformed" pattern, however, in which all the features and relations relevant to condemning the old pattern as grossly unjust continue unaltered, the only change being that the blacks are now nationals of sovereign states. And what is worse, *J* even encourages such "reform" by offering that attractive opportunity to reduce injustice without foregoing its benefits. Of course, *J* is not strictly inconsistent. But it does not make sense to let the same inequalities continue in the context of a slightly changed pattern that can so easily be reached from the old pattern through a (by the lights of *J*) merely cosmetic reform, which leaves intact all advantages, disadvantages, and correlations between them.

For all I have said, an acceptable conception of justice may allow persons to form a state, to draw borders around it, and to exclude other persons from it. It may also allow that compatriots attempt—solely among themselves—to meet a standard of fair cooperation that is much more stringent than whatever minimal constraints (of "basic justice") it may impose on cooperation in general. So, an acceptable conception of justice may allow the creation of states to lead to a higher standard and to a double standard. What it should not allow is that persons, by making others into foreigners, can dilute or evade even its minimal standard, which would otherwise govern relations between them. I say 'minimal', because I am not here worried about a conception of justice that allows a group to evade a higher standard, e.g., by seceding from a state that is more egalitarian than this conception minimally requires. My point concerns only whatever minimal standard a conception of justice applies to *any* pattern of interaction—even in the absence of communal bonds, fellow feeling, a common history, shared projects, and so forth. However such a conception may fix this standard of basic justice, it should not allow groups unilaterally to circumvent or dilute it. Specifically, it should not allow such a group, by forming a social arrangement within the larger pattern of interaction, to bring a new moral standard into play which may outweigh, trump, or cancel the minimal protections justice requires for those left out of this arrangement.[13] The introduction of states should not have that sort of moral efficacy, should not produce that sort of moral change. A conception of justice should not guide persons toward, and reward them for, deliberately creating borders in order to "just-ify" a given correlation of harms and benefits, i.e., in order to make it compatible with the requirements of basic justice without materially changing it.

Here it is possible, once again, to make a counterclaim: it is regrettable that the old white leaders, by instituting the homelands, are able morally to avoid a genuinely reformed regime under which blacks would have attained legal, political, and economic equality. But this is not regrettable *on the whole,* because *J*'s rationale also includes the view that the multiplication of states has great moral value that here outweighs the value of nonexploitation. We can leave this counterclaim aside as implausible: no such rationale is reflected in the conceptions of justice that are seriously held in our society.

This concludes the illustration of my antiloophole constraint: a code of ethics or justice must not without compelling reason encourage changes through which its own minimal standard of common decency or basic justice can be deliberately undermined through the creation of social arrangements and the special moral ties these involve.

VI AN OBJECTION

There is a way of defending the belief in the moral efficacy of creating social arrangements. *E* and *J* do not generate the offending incentive, one might claim, so long as they forbid forming or joining social arrangements in such cases.[14] In this way one can preserve the judgments that the lawyer-initiated conversion of the apartment complex is not wrong, and that economic relations in southern Africa are not unjust once the homelands have been introduced: it was wrong to create the relevant social arrangements with the motives in question, but their creation nevertheless has the intended moral effect. *E* and *J* could then hold the creators of the social arrangements—those who hire lawyers or institute national borders in cases like 1 and 2—fully responsible for the harms inflicted upon outsiders.

---

[13] Once national borders have been drawn in southern Africa, it will, with time, come to seem obvious that the white state may be organized democratically so that its government will have strong incentives vigorously to promote the interests of its citizens against the (however disadvantaged) people of neighboring states, who, after all, have their own (however impotent) governments to take care of their interests.

[14] Even if *E* and *J* are understood in this way, they could still be attacked for

I consider this defense implausible on three counts. First, the proposed moral codes would provide no reasons to derail harms that are yet in the future. Thus, suppose your father dies before the conversion is initiated and you inherit the complex with the lawyer already in charge of it. Since what is wrong is only the hiring of the lawyer (spilt milk) and not her continued efforts on your behalf, neither you nor she have any moral reason to oppose the conversion—even though it will bring about serious harms in which you both are essentially involved. You and the lawyer are not just bystanders who are merely failing to prevent harms done by another: at least one of you (and no one else) is intending to increase the rental income, deciding that and how the conversion should take place, increasing the rent to its new level, evicting the tenants who cannot pay, and profiting from the conversion. An analogous point applies in case 2.

Implausible is also, secondly, the idea of saddling the creators with open-ended posthumous responsibility. After all, the harms they are held responsible for (unlike future catastrophes caused by sloppy hazardous-waste disposal) do not arise from their conduct directly. Rather, these harms are, in case 1, consciously brought about by the lawyer who, in turn, serves at the pleasure of the current owner; and, in case 2, arise in the context of a pattern that is deliberately perpetuated by later whites.

Third, and most importantly, it is unclear how it can be so wrong for your father to create the agent-client team, given that, as the objection contends, this team does nothing wrong. And similarly for the old white leaders who initiate a pattern of interaction that, by hypothesis, is not unjust. Here it even seems that, if anything, we should have to congratulate these leaders for having replaced the

---

giving regrettable *concrete* incentives. But the complaint that they contain loopholes is now blocked, for ideal adherents of *E* and *J* would not implement either your father's idea or the homelands' reform, if doing so were forbidden.

former unjust pattern by the "reformed" one—exactly as we would have congratulated them for creating a new, just pattern that does not disadvantage blacks.

For these three reasons, we should disqualify a revision of *E* that meets the antiloophole constraint by allocating responsibility for the suffering of the tenants to the creator of the landlord-lawyer team while assigning no responsibility to this team itself. The team's aggregate wrongdoing in initiating the conversion must not be less (though it may of course be greater) than the solitary landlord's wrongdoing in initiating the same conversion singlehandedly.[15] Similarly, we should disqualify a revision of *J* that meets my antiloophole constraint by allocating responsibility for the suffering of blacks to those who internationalized the pattern while finding no fault with the "reformed" pattern itself, nor with those who maintain it in existence.

The general point is that, when matters of common decency or basic justice are at stake, a morality must not be sensitive to changes that are consciously instituted in order to secure a more favorable evaluation, but are (by the lights of this morality itself) merely cosmetic.

VII. STRENGTHENING

Let me proceed by sketching—again for the topic of social arrangements—a way of strengthening this initial thesis. Suppose, in variation of case 1, that we find new evidence about your late father's motivations. Although he might have hired the lawyer in order to enable you to reap the benefits of the conversion without wrongdoing, he actually did this because he thought you would not cope well with the requisite paperwork. Would this finding make it permissible, or more permissible, for you and the lawyer to convert the complex? Or suppose, in a similar variation of case 2, that we are ignorant of why the old white leaders (of the 1980s) instituted the homelands. Does such ignorance make it impossible to evaluate the justice of the economic relations that prevail in southern Africa more than a century later?

This would be quite strange. For the creators of the relevant social arrangements are not directly involved in the production of the harms that are currently inflicted upon tenants and blacks. Their motives thus seem irrelevant to the question what moral protections these victims should enjoy against the landlord-lawyer team or against the present whites who are upholding the multinational economic regime in southern Africa of 2100.

---

[15] Although I have discussed how responsibility should be distributed between the team and its creator, I am not considering in this essay how responsibility should be distributed within that team.

This consideration shows that it is implausible to close the loopholes by the narrowest revision of *E* and *J*, which would cancel the moral efficacy of creating (or joining) a social arrangement if and *only if* this creation was actually motivated by the offending incentive. But if information about why a social arrangement was originally created or joined is irrelevant, then the loopholes we have analyzed show more than one might have thought at first. They indicate not merely that the moralities in question are overly sensitive to certain *changes* that are, and are meant to be, merely cosmetic; but, more generally, that these moralities are overly sensitive to certain merely cosmetic *differences*. The loopholes indicate that *E* and *J* overstate not merely the moral efficacy of creating social arrangements, but, more generally, the moral significance of such arrangements themselves.

To clarify this point, let me go through the strengthening move once more in a tax-code context. Suppose a tax code contains the provision that firms whose name begins with a 'P' shall be exempt from a certain tax. Assuming that firms can choose to name or rename themselves as they please, and assuming also that no serious purpose is served by bringing the capital 'P' to social prominence, this exemption constitutes a straightforward loophole; it provides an inherently regrettable ideal reward incentive. This incentive can be eradicated by excluding from the scope of the exemption any firms that deliberately name or rename themselves in order to reduce their tax liability. But, on reflection, this revision seems implausibly narrow: Why should the tax treatment of Pacific Ice-Cream (affecting its current owners) hinge on the motives of those who, perhaps quite some time ago, chose this name for the firm? Why should it hinge on any historical facts about its naming? Inherently regrettable ideal incentives may thus indicate the need for a broader revision of the relevant code—one that eradicates the code's sensitivity to the cosmetic *difference*, rather than only the opportunity to exploit this sensitivity through cosmetic *changes*.

Hence my hypothesis: a loophole in a morality supports the conclusion that this morality is sensitive to a merely cosmetic difference. In light of case 1, we should thus question quite generally why the minimal moral concern tenants are owed should vary with whether their building is managed by a lawyer in the owner's behalf or by the owner himself. And case 2 should lead us to question quite generally why national borders should make a difference to what standards an economic regime must meet as a matter of basic justice.

We can test this hypothesis by examining two argumentative strategies that should work if this hypothesis is correct. In the next two sections, I shall try to make plausible that these two strategies do indeed have some appeal, and thereby, I hope, reinforce my hypothesis.

VIII. FICTIONAL HISTORIES

Arguments of the first strategy consist of two steps that correspond, respectively, to my initial thesis and the strengthening move. Imagine someone who believes there is nothing wrong with a lawyer's converting her client's apartment complex in his behalf, even though it would be wrong for the landlord to initiate this conversion on his own. The first step in challenging this belief consists in constructing a fictional history of the existing landlord-lawyer relationship which differs from the actual history by taking the creation of this relationship to have been motivated by the desire to exploit precisely the moral differential that our interlocutor is postulating. This fictional history may recount, for example, how the landlord's father had once hired the lawyer in order to make possible a conversion without wrongdoing. Through use of arguments from sections IV and VI, we then try to get our interlocutor to concede that an acceptable morality must not encourage that fictional conduct, nor allow it to succeed. That is, in the fictional history the hiring of the lawyer fails to enable a conversion without wrongdoing.

To transform this conclusion about a fictional case into a conclusion about the actual case, we must draw upon the strengthening move. The second step consists, then, in challenging the moral relevance of the difference between the actual and the fictional history: Why should the fact that our landlord hired the lawyer from quite harmless motives (perhaps because he wanted to live in another town) entail that he is morally entitled to reap the extra profits at the tenants' expense which he would not be morally entitled to reap in the fictional scenario? And why should this same fact entail that the tenants are not morally entitled to the preservation of their community whereas they would be otherwise? Failing a convincing reply to these questions, we are bound to conclude that the conversion with its actual history is wrong if the conversion with the fictional history is so. Both steps together show that the existence of the lawyer in the actual conversion episode cannot have the moral significance that our interlocutor had ascribed to it.

The parallel explosion of our international justice example would go like this: many are initially inclined to approve of a regime in which each state's citizens are free to elect whatever leaders they prefer and these elected leaders are expected to safeguard and promote the interests of the citizenry and the state, even at quite some cost to foreigners. People tend not to be disturbed by existing extreme international inequalities in bargaining power which engender, and are reinforced by, radical inequalities in rights and powers, in income and wealth, and in access to education and health care. Yet they do find such inequalities unjust when they occur within a single state.

To challenge such a conception of justice, we first construct a fictional history of the present world order which differs from our actual history in that in it the creation of national borders was motivated by a desire to exploit precisely the moral differential that our interlocutors are postulating. About this fictional history we draw the same conclusion as we drew about the South Africa case: national borders so created lack the alleged moral significance. Economic interaction on the basis of extreme inequality is as unjust after their creation as before.

The second step of the challenge consists in the question how the moral evaluation of the present economic relations between the First and Third Worlds can depend on the motives of those who once transformed colonies into sovereign states. How can the way in which the actual history of our global economic arrangements differs from their fictional history make a moral difference to the claims that today's rich and poor have upon global economic institutions? Why should economic relations on the basis of radical inequality be any more acceptable in our world now than they would be in the hypothetical southern Africa of 2100? Failing convincing replies to these questions, we must conclude that terms of cooperation that, violating basic justice, would be strictly unjust within *any* state must also be unjust across national borders.

IX. PUZZLES OF EQUIVALENCE

I have suggested that we should avoid inherently regrettable ideal incentives toward creating or joining social arrangements by denying altogether the moral significance of social arrangements in matters of common decency and basic justice. A further test of this proposal is provided by an argumentative strategy that can challenge an approving moral judgment directly—without any reference to actual or fictional incentives. If someone condones a course of conduct or pattern of interaction involving social arrangements, we may challenge this judgment by constructing a parallel scenario that differs from the first only in that it lacks the relevant social arrangement(s). Such a challenge may show that the judgment in question implicitly relies upon the moral significance of a difference that, on reflection, turns out to be merely cosmetic. Through this strategy one can challenge the alleged moral significance of differences even in cases where it would not give rise to inherently regrettable incentives, i.e., where it would not be practicable to exploit this significance.

Our landlord case can be recast so that it centers around the moral differential itself, rather than the incentives this differential may generate, by telling a story about two landlords, in similar circumstances, whose apartment complexes get converted at great cost to existing tenants. One of the landlords employs a lawyer, who initiates the conversion in the owner's behalf, while the other landlord is acting on her own. Why should the first conversion (no matter how the lawyer was originally hired) be any less wrong than the second?

Let me give two sample challenges in this vein to moral judgments that it would be virtually impossible to attack in the fictional history mode. Imagine someone who believes that it is ethically permissible, if not obligatory, for a lawyer defending a rapist (whose guilt is not in doubt) to use any and every legal means likely to reduce his client's sentence. Thus it is perfectly all right for him to try to divert the attention of the jury and the general public with various fabrications about the victim's sexual past and preferences which, even if true, would not have the slightest legal relevance. This judgment can be challenged by constructing a parallel scenario without attorneys, in which accuser and defendant present their cases in person. In this scenario, a man, after having raped his victim, is now trying to divert the attention of the jury and the general public with various fabrications about the victim's sexual past and preferences which, even if true, would not have the slightest legal relevance. The challenge is: if it is unethical so to slander the victim for the sake of a lighter sentence in one's own behalf, then why should it not be unethical to do so in behalf of another? What is it about the existence of attorney-client teams which supposedly overcomes the ordinary moral claims of rape victims against groundless slander for the sake of a lighter sentence?

The second example concerns an aspect of international relations. Suppose someone believes that it is permissible for the elected President of the American people, acting in the best interests of his constituency, to bring about the deaths of many unarmed civilians in order to reduce dangers and losses to American soldiers. He may, for example, order the aerial bombardment of population centers so as to demoralize the government and armed forces of the enemy and thus to increase the chances for their surrender. This moral judgment can be challenged by constructing a parallel scenario without social arrangements, such as the following: operating on their own in enemy territory, individual American soldiers cause the deaths of many unarmed civilians in order to reduce their own dangers and losses. For example, they indiscriminately throw hand grenades into civilian homes so as to protect themselves against the possibility of enemy snipers hiding inside.

When comparing these two scenarios, we must assume, of course, that other things are, as far as possible, the same. In particular, both cases are such that the following facts obtain, and are known to obtain by the relevant decision makers: the U.S. involvement in the war is justifiable. Victory or defeat, with any cosmic consequences either of these might entail, are not at stake. Interspersed among the civilians, there is a (small) proportion of enemy soldiers (the same in both cases), who get killed as well. Both scenarios produce the same predictable (and rather high) ratio of their civilians killed to our soldiers saved: there is the same unilaterally imposed correlation of benefits for us and harms deliberately inflicted upon noncombatants. The challenge is then to explain why the destruction of civilian targets as ordered by the President on behalf of his constituents should be ethically permissible while a similar destruction of civilian targets by individual soldiers is not. Why should whether the civilians of Hiroshima may be sacrificed for the sake of American soldiers hinge on whether or not the Americans make this decision individually or through an authorized representative?

X. CONCLUSION

Even if everything I have said thus far were convincing—a rather big 'if'—it would compel no substantive moral conclusions. There is, as we have seen, vulnerability to counterclaims. A defender of the moral judgments I have challenged may be able to show that a change or difference I have diagnosed as "merely cosmetic" is morally relevant after all. She may be able to show, for instance, that her morality attributes intrinsic moral value to the creation of certain social arrangements, so that it is morally important after all that lawyers be hired, states be multiplied, and so forth.

Even if the defender can be forced to concede that the change or difference in question really is merely cosmetic, the revision of her morality might still go either way: my challenge may open her eyes to wrongs and injustices suffered by those excluded from the more powerful social arrangements that tie together the rich and mighty. But then it might, instead, also lead her to change her morality in ways that would make their suffering acceptable even in the absence of any social arrangements whatever. Thus, she might cease to find wrong even the landlord's direct destruction of his tenants' community for the sake of extra profits. She might now condone apartheid even within a single state. She might come to endorse weaker ethical constraints even upon the investment of one's own money. She might now tolerate the deliberate killing of civilians in war even when this is done by individual soldiers acting on their own. And she might convert to a libertarian conception of economic justice that permits radical inequalities even among compatriots. I hope, of course, that no one will be attracted to such revisions. But I cannot rule them out. The examples show then that the incoherences we have diagnosed through the analysis of loopholes can be repaired through quite diverse revisions.

This problem highlights that I have been arguing for a *structural* constraint upon moral conceptions: I am trying to upset received moral convictions not by confronting them with a new set of substantive moral intuitions, but by trying to show how we sometimes structure and organize our moral convictions in a way that tends to obstruct and undermine their practical significance.

The character of such structural defects in a morality is likely to bear some relation to the position of those holding it. For people in the developed West, for example, it is comfortable to believe that—thanks to intervening borders—the moral demands on our relations with people in Latin America, say, are much weaker than even the minimal moral demands on relations with one's compatriots. If this is right, the present essay may also make a contribution to the explanation of complacency: How can we live so easily with the enormous advantages we enjoy in our political and economic relations with Third-world populations?

My main objective, however, was to sketch a way in which we might advance our (ordinary and academic) moral discourse. Even while there is a good deal of substantive disagreement, it may still be possible to establish some structural constraints that any acceptable (conception of) morality must satisfy. Agreement on such constraints would tend to narrow the existing spectrum of moral debate. It would thereby make a contribution toward moral convergence, which cannot but enhance the practical significance of moral values.

THOMAS W. POGGE

Columbia University

BOOK REVIEWS

Kant's Theory of Freedom. HENRY E. ALLISON. New York: Cambridge University Press, 1990. xii + 304 p. Cloth $49.50, paper $15.95.

Kant held that human beings possess transcendental freedom as well as merely practical freedom. That is, he held not merely that human action is capable of being governed by principles as opposed to being determined solely by the occurrence of sensory desires and impulses, but also that such governance of our action by principles is incompatible with the determinism that he took to rule the empirical world without exception, so that in any case of action we could always have chosen to do otherwise than we did. The two claims are not identical: one can hold that it is merely a fact of nature that can itself be given a causal explanation that human behavior is at least sometimes controlled by adherence to certain principles as opposed to the mere presence of certain inclinations. If this is so, then the existence of practical freedom does not by itself entail the existence of transcendental freedom. It is customarily supposed that Kant thought that the *possibility* of transcendental freedom could be established by theoretical considerations alone, but that the *reality* of transcendental freedom could be established only by appeal not merely to practical reason in general but specifically to the conditions of moral imputation, which requires the possibility that human beings could always have chosen to do otherwise than they in fact did as a necessary condition for the imputation of praise and blame.

In his new book, Henry Allison rejects this view, arguing that Kant thought that the reality of transcendental freedom is a necessary condition of the possibility of rational agency in general, not just moral agency—that is, a condition of the possibility of action in accordance with any kind of principle, not just the principle of morality—and, what is more, that Kant's basis for this view of agency is essentially theoretical rather than practical. This position accounts for what would otherwise be the counterintuitive organization of Allison's book, namely, that he begins by offering the argument for the reality of transcendental freedom and the interpretation of transcendental idealism that he holds to be necessary to make sense of transcendental freedom (in part I) before discussing the fundamental principle of morality that would ordinarily be thought to motivate Kant's belief in transcendental freedom (in part III). (This organization can also be explained by the chronological fact that Kant introduces the theory of transcendental freedom in the *Critique of Pure Reason* before attempting the detailed connection of freedom to the fundamental principle of morality in the *Groundwork of the Metaphysics of Morals* and *Critique of Practical Reason.*)

Allison's methodological presuppositions are very different from those I would bring to the study of Kant's various discussions of freedom. Allison, in this regard certainly like many other distinguished commentators, is deeply committed to the assumption that the views of a great philosopher like Kant must possess a profound unity that can be brought out by a sympathetic interpretation. My own studies of Kant have persuaded me that Kant's greatness lies more in some of his particular analyses and arguments and in his recognition of the complexity of the connections among them than in his pretensions to systematicity. In particular, I believe that the appearance that the three critiques constitute a single, consistent system, an appearance which Kant certainly cultivated, is deeply illusory; that in fact Kant had no thought of needing a second critique when he wrote the first nor of needing a third when he wrote the second; and that his views underwent a great deal of evolution and revision over the course of these three great but to some degree independent works. For this reason, my own interpretations of Kant have placed me in the more critical and revisionist camp of commentators often reviled as "patchwork" theorists.

But it is my view that Allison's apparently attractive methodological assumptions nevertheless have damaging consequences for the present work. Part of Allison's project is to show how the single conception of the transcendental idealism of the *Critique of Pure Reason* which he expounded in his 1983 book *Kant's Transcendental Idealism: An Interpretation and Defense*[1] provides the framework necessary for understanding Kant's theory of freedom, an exercise that is required because this particular interpretation of transcendental idealism seems singularly ill-suited to this task. Here I believe that Allison has compounded the difficulty of his task by his refusal to recognize that the interpretation of transcendental idealism which he prefers, although it is certainly suggested by some of Kant's remarks, is by no means the only one for which Kant argues and may not be the one best suited for understanding the assumptions that Kant makes when it comes to the metaphysics of moral philosophy. But beyond this, Allison's overarching project of unification in the present book, which is nothing less than the aim of finding the foundation of Kant's moral argument for transcendental freedom in a single conception of agency underlying his theoretical as well as practical philosophy, puts the value of his work at great risk if this underlying conception of agency, whether or not Kant gives it the degree of importance that Allison does, is not very persuasive. In my view, it is not, and therefore the value of interpreting Kant's moral theory of freedom in light of an exclusive emphasis on this one idea is dubious.

---

[1] (New Haven: Yale, 1983), esp. chs. 1 and 2, pp. 3–34 See my critique in *Kant and the Claims of Knowledge* (New York· Cambridge, 1987), ch. 15, pp. 333–44.

My main criticism of Allison's book, therefore, is this. The theoretical proof of the reality of transcendental freedom which he finds in the *Critique of Pure Reason* may well be suggested there. But it sits uneasily with Kant's claim to have shown no more than that "freedom is at least *not incompatible* with nature" (A 558/B 586), and is in any case entirely inadequately expounded and defended. Further, although Kant may not yet have recognized the failure of this approach in 1785 when he published the *Groundwork*, he clearly did by 1788 when he argued in the *Critique of Practical Reason* that the argument for the reality of transcendental freedom must be strictly practical, not theoretical. Allison's treatment of the two major ethical treatises in part III of his book might appear to recognize this fact, for it looks as if he there falls back into the more traditional search for an argument for the reality of freedom from moral considerations alone. But in fact he tries to avoid the traditional approach, and instead reinterprets the central argument of the *Critique of Practical Reason* as turning on the conception of transcendental freedom as the necessary condition of any kind of agency which he has found in the *Critique of Pure Reason*. In my view this conception is too weak to carry the weight Allison assigns to it, and it is precisely because he knew this that Kant did not in fact deploy this conception in the second *Critique*. Allison thus undermines any natural reading of Kant's progress in his moral theory.

I

Allison's interpretation is organized around three theses. The first is what he calls, following the lead of *Religion within the Limits of Reason Alone*, Kant's *incorporation thesis*, the view that no impulse or desire ever motivates human action by the mere psychological fact of its occurrence but does so only by the agent's taking it up—incorporating it—into a maxim or principle that permits or endorses action on that desire (5–6, 39–40). The second thesis, which Allison does not name but which following the model of the first may be called the *spontaneity thesis*, is the claim that the act of adopting a principle, whether moral or not, as well as that of incorporating a desire into one's ends on the basis of such a principle, cannot be understood as a deterministic effect of some antecedent cause but can only be understood as a manifestation of transcendental freedom. These two theses constitute the gist of Allison's interpretation of Kant's "thick" conception of rational agency in general. The third thesis, which he entitles the *reciprocity thesis* (201–13), is that the ability to conform one's behavior to the moral law requires transcendental freedom but also that the possession of transcendental freedom entails the recognition of the validity of the moral law. This thesis, which is certainly asserted in both the *Groundwork* and the *Critique of Practical Reason*, is deployed by Allison in an attempt to prove that Kant thought that our possession of rational agency in general, which suffices to prove our transcendental freedom, also entails our recognition of the validity of the moral law in particular.

There can be no doubt that Kant subscribes to the incorporation thesis (as others have also shown),[2] and in part II of his book Allison shows its importance for understanding a number of important issues in Kant's moral psychology. These include the problem of acting from the motive of duty, where Allison resolves the dispute over whether such action requires the absence or inefficacy of inclinations by showing that the issue is never the mere presence of inclinations but whether they have been taken up into a maxim that is licensed by the moral law or one that is not (chapter 6); Kant's theory of moral disposition and radical evil (chapters 7 and 8); and his response to Friedrich Schiller's attempt to remove the divide between moral sentiment and moral principle (chapter 10), where again the issue can never be merely what kinds of sentiments one has but how they are incorporated into one's maxims. But Allison also defends the metaphysical claim that the incorporation thesis implies the spontaneity thesis, that is, that action on principles and never merely *per stimulos* alone requires transcendental freedom. Here his account is unconvincing.

The argument that Allison finds in Kant is just this: since rational agency requires incorporating or "taking up" into one's principles whatever inclinations may present themselves to one, or else rejecting action on such inclinations because of their conflict with these principles, regardless of whether these are moral or merely prudential (36), rather than merely react to their presence, such incorporation is an act of spontaneity rather than mere receptivity, just as using the faculty of understanding to conceptualize and judge data that are passively received by the senses is also an act of spontaneity rather than receptivity; and such acts of spontaneity cannot be included within the realm of causally determined events (36–7). Thus, such acts must be ascribed to the "intelligible" rather than merely "empirical" character of the agent. In Allison's words, "to conceive of oneself (or someone else) as a rational agent is to adopt a model of deliberative rationality in terms of which choice involves both a taking as and a framing or positing. Since these activities, as expressions of spontaneity, are themselves merely intelligible (they can be thought but not experienced), it is necessary to attribute an intelligible character to the acting subject" (38). Because of their spontaneity, any acts seen as products of practical freedom (action according to a principle) must also be seen as products of transcendental freedom (action free of determinism).

---

[2] A clear account of it, though without Allison's title, may be found in Allen W. Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, Allen W. Wood, ed. (Ithaca: Cornell, 1984), esp. pp. 78–80.

Kant may well suggest such a position in the first *Critique* (see A 546–8/B 574–6). Even so, it is clearly a poor argument. Conceptualization and judgment may well be "spontaneous" rather than "receptive" in that they are not products of sensory stimulation alone but emerge only from internal cognitive processing of sensory stimuli; but that does not imply that there is anything nonnatural or nondeterministic about such cognitive processing. It may just be a fact of nature like any other that we process our sensory stimuli the way we do, even if no other creatures in nature do. Likewise, although incorporating a desire into one's maxims is "active" in a way that the mere occurrence of the desire itself is not, it hardly follows that this "activity" itself cannot be explained within a naturalistic framework. Without a great deal more argument, the contrast between "spontaneity" and "passivity" is only an unredeemed metaphor that cannot support an inference to the real existence or even the mere possibility of a breach with determinism. But this contrast is all that Allison's defense of the spontaneity thesis amounts to; he does not attempt to prove that mental "activity" is really incompatible with determinism.

There are also problems with Allison's attempt to defend Kant's reconciliation of this account of intelligible character with the determinism that Kant supposes to prevail in the empirical realm by the use of his well-known "two-aspect" rather than "two-world" interpretation of transcendental idealism. On this account, appearances, where determinism rules, and things in themselves, where spontaneity is possible, are not to be understood as two different sets of objects, but rather just as a single realm of objects represented through two different conceptions or points of view: the concept of appearances, which includes those features necessary for us to represent objects ("epistemic conditions"), and the concept of things in themselves, which excludes them. Because Kant's argument for universal causation shows it to be the necessary condition for time determination, but time is only an epistemic condition and therefore a feature of our conception of appearances but not a feature of things in themselves, there is supposed to be "conceptual space" for the conception of ourselves as things in themselves who possess transcendental freedom (44). No disturbing metaphysics but only the comfortable conceptual distinction between two points of view is therefore necessary to justify Kant's assertion of the compatibility of empirical determinism and transcendental freedom.

Kant doubtless suggests such an interpretation of his transcendental idealism at various points in the first *Critique* (e.g., B 307), but there are also places where his arguments actually imply the stronger view, not merely that our *concepts* of things in themselves exclude space and time but that the things themselves *are not* spatial and temporal (e.g., A 26/B 42). It would seem to be the latter rather than former interpretation that is necessary to make sense of his claims about noumenal freedom. In any case, Allison's defense of his interpretation as the basis for a defense of Kant's theory of transcendental freedom does not work.

First, when he defends Kant's ascription of freedom rather than determinism to our "intelligible character" by saying that intelligible character is only thought but not experienced (see also 52), Allison ignores one of the historically most significant aspects of Kant's treatment of causation: Kant's reply to Hume's doubts about the meaningfulness of our idea of causation (as opposed to those about the justifiability of induction) depends precisely on his recognition that the idea of necessary connection is not to be sought among our impressions, but is rather a category brought to our impressions and applied to them through a schema. In other words, causation and thus determinism are also *thought* and not merely experienced—looking for causation in mere experience was what led Hume to despair.[3] Thus one can hardly show that transcendental freedom and determinism are compatible merely by distinguishing between thought and (mere) experience.

The problems with Allison's version of Kant's compatibilism go deeper than this, however. On the face of it, merely distinguishing between a point of view on objects that includes determinism or the ground for it and one that excludes it cannot open up "conceptual space" for transcendental freedom, because merely abstracting from a particular property of an object in one of our concepts of it does not mean that the object does not in fact have that property. As Terence Irwin[4] has written: "Now if an event is determined, it is true of it under all descriptions that it is determined, even though only some true descriptions . . . show why it is determined" (*ibid.*, p. 38). This is the most obvious objection to the use of the "two-aspect" interpretation of transcendental idealism for a theory of freedom. Allison tries to defend himself against Irwin's formulation of the objection by arguing that it "rests on the assumption that being causally determined is a necessary or essential property of any occurrence of which it is predicated," and then arguing that for Kant "the concept of causality is merely an epistemic condition, more specifically, a condition of the representation of objective temporal change" (44), which he presumably therefore considers to be merely an accidental or inessential property of objects. But this either misses the point or else in fact gives up the two-aspect interpretation of transcendental idealism. It misses the point if it supposes that Irwin's objection depends on the assumption that determinism is *necessarily true* of human agents in all possible worlds—Irwin's point is simply that, if determinism is in fact *true at all* of humans as they actually are, then it remains true of them no matter how they are described, for leaving determinism out of some description *does not make it false*. Adopting Allison's apparent alternative of viewing our behavior as only "accidentally" rather than "essentially" deterministic does not make it nondeterministic. (Maybe it makes the behavior of our "counterparts" in some other "possible world" nondeterministic—but what would that prove about *us?*) Alternatively, if Allison is responding that deterministic causality is only a feature of our *representation* of ourselves in our empirical surroundings but *not* a feature of ourselves as we really are, because "epistemic conditions" are necessarily merely subjective and therefore *not true* of things in themselves, then what becomes of the difference between the "two-aspect" and "two-world" interpretations of transcendental idealism? Allison is forced to maintain that determinism must be *true* of our representations of ourselves but *false* or at least possibly false of ourselves as we really are, where our representations and ourselves cannot be the same objects—which is all that the "two-world" interpretation has ever maintained. So Allison cannot maintain his interpretation of transcendental idealism yet answer Irwin's objection.

---

[3] Or so Kant thought. Actually, Hume himself was satisfied with his account of the origin of the idea of necessary connection in an internal sensory impression in *Treatise*, bk. I, pt. III, and only fell into despair when he could not find a sensory impression to serve as the origin of the idea of distinct existence in pt. IV.

---

[4] "Morality and Personality: Kant and Green," in *Self and Nature in Kant's Philosophy*.

Finally, Allison's argument that his interpretation of the incorporation thesis and intelligible character frees us from traditional worries about the necessity of acts of choice that must be timeless because they are acts of noumenal rather than phenomenal agents is equally inadequate. His position is that because incentives "do not motivate by themselves causing action but rather by being taken as reasons and incorporated into maxims" in accord with principles supplied by reason, where reason "is not literally an efficient cause" (51) and the "act of incorporation," "as an expression of spontaneity, cannot be regarded as the causal consequence of antecedent conditions" (52), we do not have to think of the incorporation of desires into maxims in temporal terms at all, thus as either in time or timeless—we are simply in the space of reasons, not of temporal events. But this is clearly false. While the fact that adherence to a certain principle *justifies* or *excludes* action on a certain incentive may not itself be a temporal fact, the fact that a particular agent *holds* a certain principle and *has justified* or *excluded* action on a given incentive on the basis of holding this principle is just as temporal a fact as the fact that a certain desire occurs in the agent. Holding a principle may be a disposition rather than an event, but it is still temporal, as is obvious from the fact that one can start or cease to hold a principle, as in deciding to adopt it or surrender it. And of course, *applying* or *deciding* to apply a certain principle to a certain desire may be as much of a temporal event as the occurrence of the desire itself. Even if it is not done self-consciously and its temporal location cannot be pin-pointed, it is still just as much a temporal event as the occurrence of the desire, which may not be a conscious and precisely datable event either. So holding that as rational agents we are never moved to action merely by the mere occurrence of incentives but rather by the incorporation of such incentives into our maxims or principles hardly removes our rational agency from the sphere of temporal discourse by itself. We must either accept this result, whatever its implications for freedom, or defend the unpleasant doctrine of timeless events of noumenal choice.

I conclude, then, that although Allison may be right that Kant attempted to derive transcendental freedom from the spontaneity of thought in general in the *Critique of Pure Reason*, this attempt was an utter failure; and Allison's own attempt to defend Kant's account of transcendental freedom with his "two-aspect" interpretation of transcendental idealism does not fare much better. Since Allison's interpretation of Kant's ultimate derivation of transcendental freedom from moral considerations in part III of his book rests on this theoretical argument offered in part I, most of that discussion also collapses with the failure of the latter.

II

Part III discusses the connections between transcendental freedom and the fundamental principle of morality. Allison might seem to fall back here into the traditional interpretation that Kant intends the actuality of freedom to be derived from purely moral considerations, but the purpose of his discussion is in fact to show that Kant's supposedly moral argument for freedom depends not on traditional worries about imputation but on the spontaneity thesis.

Allison's discussion is structured by the customary assumption that Kant first intends to establish an analytical biconditional between freedom and the moral law, and then a further synthetic proposition, either, as in the *Groundwork,* that we are indeed free and therefore bound by the moral law, or, as in the *Critique of Practical Reason,* that we are indeed bound by the moral law and therefore truly free. In either case, the argument begins with what Allison calls the reciprocity thesis, that is, Kant's claim that only a free will could be capable of acting in accordance with a purely formal principle such as the categorical imperative but also that a free will could only act in accordance with such a formal principle (see 202, where Allison cites *Groundwork* 4:446–7 and *Practical Reason* §§5 and 6).

Allison treats the first conjunct of the reciprocity thesis, the claim that conformity to a formal principle requires a free will, as nonproblematic, although that is not obviously the case if the freedom of the will is taken to be transcendental freedom. Conformity to a formal law, at least under some circumstances, could be explained by mere practical freedom, that is, the possibility of the naturalistic determination of action by a principle rather than mere inclination, unless one either assumes that the only possible empirical determinants of action *are* mere inclinations—a view that Allen Wood ascribes to Kant and Allison unpersuasively rejects (48–51)—in which case practical freedom is an empirical impossibility, or else argues that practical freedom entails transcendental freedom, which Allison has done in part I, with what success we have already seen. In any case, it is the second conjunct of the reciprocity thesis, the claim that possession of freedom (again assumed to be transcendental freedom) entails the validity of the moral law, which draws Allison's attention.

The most obvious problem with this proposition is that, if the moral law is actually taken to describe the operative causality of a free will, then an agent who fails to conform to the moral law must lack a free will and thus not be responsible for his moral failure.[5] The equally natural response to this problem is to say that Kant only means to argue that freedom of the will entails a capacity to recognize the validity of the moral law; but then this does not seem strong enough for his purposes, at least in the *Groundwork,* where he apparently wants to prove to us that we *are bound* by the moral law, not just capable of recognizing its validity, in virtue of our freedom. So Allison tries to steer a middle course, and interprets Kant's thesis as equivalent to the claim that insofar as we are transcendentally free rational agents we necessarily recognize the validity of "criteria of reasonableness" (204), or require of our maxims "legitimacy for all rational beings in similar circumstances," from which it would follow that "rational agents cannot reject the universalizability test without, at the same time, denying their rationality" (205). This might seem open to the objection that a rational agent, in particular a rational egoist, could recognize that all others "ought" to act on his own maxims without in fact having reason to will that they do so (206) (the rational egoist of course hopes that others will be irrational), but, Allison claims, this objection depends on a failure to recognize the depth of Kant's conception of rational agency. He then appeals to his interpretation of Kant's "thick" conception of rational agency to claim that a rational agent is never moved by impulse alone, but always incorporates his impulses into his maxims in virtue of a principle, and is thus, apparently, necessarily committed to the recognition of a requirement of justifiability for his principles (208). Here, however, his argument is obscure: it is just not obvious why the fact that a rational agent must always incorporate his desires into *some* principle entails that the latter must be a *universally justifiable* principle. Allison is presumably not assuming that a transcendentally free agent is just a pure reason and does not actually *have* any incentives that could *conflict* with universally justifiable principles, which would depend on a metaphysical conception of the purely rational subject and lay Allison open to the standard Hegelian objections to that. But he does not offer any other explanation of why the necessity of acting on principles implies the necessity of acting on a universally justifiable principle.

Allison's defense of the reciprocity thesis is thus unconvincing. His criticism of Kant's attempt to exploit the thesis in the *Groundwork* by providing a speculative proof that we have a transcendentally free will, which would therefore imply that we are bound by the moral law, turning upon the alleged theoretical fact of our membership in the intelligible world, is basically sound (although it does follow in tracks already well-laid by commentators such as Dieter Henrich and Karl Ameriks). But Allison insists on interpreting the argument from the moral law to freedom with which the *Critique of Practical Reason* replaced the *Groundwork's* argument from freedom to the moral law in light of his spontaneity as well as incorporation thesis, and here his story becomes problematic again.

---

[5] This objection of course goes back at least as far as Henry Sidgwick, although Allison associates it particularly with Robert Paul Wolff (95, 263 n. 35).

Allison does refer in passing to Kant's assumption that moral imputation requires freedom (e.g. 241), but on his account this is not the basis of Kant's argument. Instead, he proceeds thus. First, he interprets the conception of the "fact of reason" that Kant introduces into the second *Critique* as "the consciousness of standing under the moral law and the recognition of this law by every natural human reason as the supreme law of its will" (233). This seems correct. He then holds that this "very consciousness of the moral law as binding produces an interest" (239), but this leaves the problem of showing "that we also have the capacity to fulfil [the] requirements" of such an interest (240). On the face of it, that is, it might seem that we could have some disposition in behalf of the moral law but not sufficient freedom or strength of will to comply with it. To avoid this problem, Allison again has Kant appeal to his incorporation thesis as well as his spontaneity thesis: an interest, he says, like a maxim but opposed to a mere inclination, is necessarily a product of an act of practical reason, thus it is necessarily a product of spontaneity, which in turn entails not only the positive freedom of conducting our behavior according to a law but also the negative freedom of being free from all determinism (241). Such negative freedom is apparently sufficient to ensure that we always have the freedom to resist whatever blandishments might stand in the way of our adherence to the moral law which we recognize in the fact of reason. Thus the mere fact of our interest in the moral law proves that we have the freedom to comply with it.

This argument rests on an obscure notion of interest and the unexplained assumption that we could have clear evidence that we have such an interest in morality as opposed to a mere inclination in its behalf, which would not of course trigger Allison's spontaneity thesis. But in any case I do not find any evidence that Kant suggests such an account of interest in the second *Critique*. It seems to me clear that Kant's argument for the actuality of the freedom of the will rests solely on his belief that it is a necessary condition for the imputation of praise and blame: "the fact of reason" makes it manifest to each of us that we always ought to conform our behavior to the moral law, even when threatened with the gallows for so doing, and we judge, "therefore, that one can do something because he knows that one ought, and one recognizes that he is free—a fact which, without the moral law, would have remained unknown to him" (5: 30). There is no suggestion here of any "interest" in morality intervening between our recognition of our obligation under the moral law and our realization of our freedom to live up to it, or any hint that the latter depends upon a theoretical explanation of the nature of an interest. Kant just assumes that moral imputation—presumably not only blame for vicious deeds but also praise for virtuous ones—would be inappropriate if the subject of imputation could not in fact have chosen to do otherwise than he did. This assumption of course needs examination, but Allison does not supply it.

In the end, therefore, I do not find Allison's account of Kant's theory of freedom enlightening or compelling. It begins with a dubious theoretical proof of freedom, appears to return to the more customary moral proof of freedom, but then obscurely and unsatisfactorily reinterprets that moral proof by means of the original theoretical proof. In all of his work on Kant, Allison has adopted the stance of the sympathetic defender of Kant rather than the critical revisionist. But in the present book, it seems to me, he plays the role of revisionist himself, though unsuccessfully, by pinning his case for the underlying continuity of Kant's thought on a shaky argument from the first *Critique* which there is little evidence to suggest that Kant retained in the second. This is no advance in the interpretation of Kant's moral philosophy. We still await careful scrutiny of what was clearly the fundamental assumption of Kant's final, practical proof of freedom, the assumption that moral imputation presupposes transcendental freedom of the will, as well an answer to the equally fundamental question of whether Kant's normative thesis of the unconditional value of human autonomy requires a metaphysical thesis of the freedom of the will at all.

PAUL GUYER

University of Pennsylvania

Cambridge University Press

Cambridge Companions to Philosophy

The *Cambridge Companions to Philosophy* provide comprehensive, up-to-date, yet accessible critical analyses of major Western philosophers. Each volume contains specially commissioned essays, by an international team of scholars, which focus on central topics and texts from the work of a given philosopher, together with a substantial bibliography. New readers and nonspecialists will find these volumes the most convenient and authoritative guides currently available, while advanced students and specialists will find in each volume a conspectus of recent developments in critical interpretation. One underlying aim of the *Companions* is to dispel the intimidation nonspecialists often feel when faced with the work of a difficult, challenging thinker. The first three volumes are devoted to Kant, Marx and Freud. *Companions* on Aquinas, Aristotle, Descartes, Foucault, Hegel, Heidegger, Hume, Husserl, Leibniz, Locke, Mill, Nietzsche, Plato, Sartre, Spinoza and Wittgenstein are forthcoming.

The Cambridge Companion to Kant

Paul Guyer, Editor

This volume is the only systematic and comprehensive account of the full range of Kant's writings available, and the first major overview of his work to be published in more than a dozen years. An internationally recognized team of Kant scholars explores Kant's conceptual revolution in epistemology, metaphysics, philosophy of science, moral and political philosophy, aesthetics, and the philosophy of religion. The volume also traces the historical origins and consequences of Kant's work

Contributors. *Frederick C Beiser, Charles Parsons, J. Michael Young, Paul Guyer, Michael Friedman, Gary Hatfield, Thomas E. Wartenberg, Karl Ameriks, Onora O'Neill, J. B Schneewind, Wolfgang Kersting, Eva Schaper, Allen W Wood, George di Giovanni*

36587-2 Hardcover $69 95
36768-9 Paper $17 96

The Cambridge Companion to Marx

Terrell Carver, Editor

Marx was a highly original and polymathic thinker, unhampered by disciplinary boundaries, whose intellectual influence has been enormous. Yet in the wake of the collapse of Marxism-Leninism in Eastern Europe the question arises as to how important his work really is for us now. An important dimension of this volume is to place Marx's writings in their historical context and to separate what he actually said from what others (in particular, Engels) interpreted him as saying. Informed by current debates and new perspectives, this volume provides a comprehensive coverage of all the major areas to which Marx made significant contributions.

Contributors *William Adams, Terence Ball, Terrell Carver, James Farr, Alan Gilbert, Jeff Hearn, Susan Himmelweit, Scott Meikle, Richard W Miller, Jeffrey Reiman, Paul Thomas, Denys Turner, Lawrence Wilde*

36625-9 Hardcover $69 95
36694-1 Paper $17 95

The Cambridge Companion to Freud

Jerome Neu, Editor

Does Freud still have something to teach us? The premise of this volume is that he most certainly does. Approaching Freud from not only the philosophical but also historical, psychoanalytical, anthropological, and sociological perspectives, the contributors show us how Freud gave us a new and powerful way to think about human thought and action. They consider the context of Freud's thought and the structure of his arguments to reveal how he made sense of ranges of experience generally neglected or misunderstood. All the central topics of Freud's work from sexuality and neurosis to morality, art, and culture are covered

Contributors *Carl Schorske, Gerald Izenberg, Clark Glymour, James Hopkins, Sebastian Gardner, Bennett Simon, Rachel Blass, Jerome Neu, Jennifer Church, Nancy J Chodorow, Richard Wollheim, Robert A Paul, John Deigh, David Sachs*

37424-3 Hardcover $59 95
37779-X Paper $17 96

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS

40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.

MasterCard/VISA accepted. Prices subject to change.

To See the Buddha

What does it mean to see the Buddha?

In the 7th Century A.D. Chinese Buddhist monk Hsuan-tsang made a famous pilgrimage to India to find the heart and mind of Buddhism. Now, in a work of sparkling scholarship, philosopher David Eckel takes us on a contemporary quest to rediscover the Buddha in Indian art, philosophy, religion and literature.

HarperSanFrancisco
A Division of HarperCollinsPublishers
Also Available from HarperCollins Canada Ltd

Hardcover • $24.95

MALCOLM DAVID ECKEL

Ohio University Press

Word and Object in Husserl, Frege, and Russell
The Roots of Twentieth-Century Philosophy
by Claire Ortiz Hill
Claire Hill analyzes Frege's, Husserl's, and Russell's work and uses linguistic analyses to look again at the principal concepts of analytic philosophy.
SERIES IN CONTINENTAL THOUGHT, VOL. 17
230 pp. $34.95

The American Moralist
On Law, Ethics, and Government
by George Anastaplo
The essays collected here range from a discussion of the despair of the Cold War and Vietnam in 1966 to reflections on the euphoria over the ending of the Cold War in Eastern Europe in 1990.
640 pp. cloth $100.00

MIND MATTERS

A new series from St. Martin's Press

Series Editor: Judith Hughes

The books in the new *Mind Matters* series explore philosophical issues in a lively, jargon-free fashion written specifically for the general reader.

Can't We Make Moral Judgments?

Mary Midgley

In a lively and approachable discussion, Midgley turns a spotlight on the popular view that we no longer need or use moral judgments. She shows how the queston of whether or not we can make moral judgments must inevitably affect our attitudes not only to the law and its institutions, but also to events that occur in our daily lives.

1991 178 pp. $24.95
ISBN 0-312-06129-3

Does God Exist?

Mark Corner

Can God's existence be proved? And even if it can, is God necessarily good? Corner provides a provoking and readable introduction to the main issues of theism and atheism. With careful avoidance of theological jargon, he takes a fresh look at the question that has always been at the very roots of philosophy.

1991 180 pp. $29.95
ISBN 0-312-06130-7

Also Available in the Mind Matters Series

Reasonable Care

Grant Gillett
1990 169 pp. $29.95
ISBN 0-312-04002-4

Before Eureka

The Presocratics and Their Science
Robin Waterfield
1990 136 pp. $29.95
ISBN 0-312-04001-6

Mad or Bad?

Michael Bavidge
1989 150 pp. $24.95
ISBN 0-312-03143-2

St. Martin's Press

Scholarly and Reference Division ✦ 175 Fifth Avenue New York, NY 10010

The path out
of our era's
moral confusion

MORTIMER
J. ADLER

DESIRES
RIGHT & WRONG
THE ETHICS OF ENOUGH

Fifty years ago Mortimer J. Adler critiqued the morality of an age in his book *A Dialectic of Morals*. Now, he crowns his efforts to present Aristotelian moral philosophy to a contemporary audience in *Desires Right & Wrong*.

Dr. Adler explores the moral digressions of lust, gluttony, greed, inebriation, and pride, as well as the moral virtues of love, friendship, liberty, and wisdom. Beginning with an explanation of why man, the only animal with moral problems, is also the only animal destructive of his own life and his planet, Dr. Adler also touches on what he believes are fundamental errors when applied as "problem-solvers"—the idealism of Plato, the Stoics, and Kant; the utilitarianism of John Stuart Mill; and the realism exemplified by John Dewey.

At all bookstores or call 1-800-323-7445 to place your credit card order

MACMILLAN PUBLISHING COMPANY
866 Third Avenue, New York, NY 10022

Harcourt Brace Jovanovich College Publishers
Adventures for the Mind

Introduction to Philosophy

New edition '92

Twenty Questions: An Introduction to Philosophy, 2/e
G. Lee Bowie, *Mount Holyoke College*
Meredith W. Michaels, *Hampshire College*
Robert C. Solomon, *University of Texas at Austin*
0-15-592389-7 1992 Paper 750 pp.

Introducing Philosophy: A Text with Integrated Readings, 4/e
Robert C. Solomon, *University of Texas at Austin*
0-15-541561-1 1989 Paper 784 pp.

The Big Questions: A Short Introduction to Philosophy, 3/e
Robert C. Solomon, *University of Texas at Austin*
0-15-505412-0 1990 Paper 331 pp.

Philosophy: An Introduction to the Art of Wondering, 5/e
James L. Christian, *Rancho Santiago College*
0-03-030414-8 1990 Hardcover 608 pp.

Logic and Critical Thinking

New edition '92

Practical Logic: An Antidote for Uncritical Thinking, 4/e
Douglas A. Soccio, *Shasta College*
Vincent E. Barry, *Bakersfield College*
0-03-073907-1 1992 Paper 512 pp.

New '92

The Critical Edge: Critical Thinking for Reading and Writing
Vincent E. Barry, *Bakersfield College*
0-03-047522-8 1992 Paper 602 pp.

Introduction to Logic and Critical Thinking, 2/e
Merilee H. Salmon, *University of Pittsburgh*
0-15-543062-9 1989 Hardcover 393 pp.

Ethics

New '92

Thirteen Questions In Ethics
G. Lee Bowie, *Mt. Holyoke College*
Kathleen Higgins, *University of Texas at Austin*
Meredith W. Michaels, *Hampshire College*
0-15-591744-7 1992 Paper 650pp

New '92

The Moral Life
Steven Luper-Foy, *Trinity University*
Curtis Brown, *Trinity University*
0-03-033969-3 1992 Paper 480 pp

New '92

Ethics: Problems and Principles
John Martin Fischer, *University of California, Riverside*
Mark Ravizza, *University of California, Riverside*
0-03-047527-9 1992 Paper 366 pp

New '92

Corporations in the Moral Community
Peter A. French, *Trinity University*
Jeffrey Nesteruk, *Rider College*
David T. Risser, *Millersville University*
with John Abbarno, *D'Youville College*
0-03-030782-1 1992 Paper 176 pp.

New '92

Values and Public Policy
Claudia Mills, *University of Maryland, Baltimore*
0-15-594711-7 1992 Paper 573 pp.

New '92

To Believe or Not to Believe: Readings in the Philosophy of Religion
E. D. Klemke, *Iowa State University*
0-15-592149-5 1992 Paper 800 pp

College textbooks published by Holt, Rinehart & Winston, Inc., are now published by Harcourt Brace Jovanovich College Publishers

HBJ

Harcourt Brace Jovanovich College Publishers
College Sales Office
7555 Caldwell Avenue
Chicago, IL 60648
1-708-647-8822

New & Noteworthy

CREATIVE VENTURES

By PAUL WEISS. Foreword by GEORGE KIMBALL PLOCHMANN. The author systematically maps creativity in its many manifestations — in the arts, in mathematics and the sciences, in moral development, in social movements, and in government — tracing the combination of factors from which a creative work arises. (1729-X) $45.00

JOHN HENRY NEWMAN
Reason, Rhetoric and Romanticism

Edited by DAVID NICHOLLS and FERGUS KERR, OP. Offering a more challenging critical assessment than has been previously available, the authors explore the life and thought of the eminent Victorian philosopher and theologian. (1758-3) $27.50

SOURCES OF SEMIOTIC
Readings with Commentary from Antiquity to the Present

By D. S. CLARKE, Jr. Furnishing both an essential selection of original writings and a concise interpretive history, Clarke deals with signs in the early Greek period as well as the medieval, post-Cartesian, and modern eras.
(1613-7) $29.95 cloth / (1614-5) $14.95 paper

INDEX [TO THE COLLECTED WORKS OF JOHN DEWEY, 1882-1953]

Edited by ANNE S. SHARPE. Associate Editor, HARRIET FURST SIMON. Assistant Editor, BARBARA LEVINE. This cumulative index to the thirty-seven volumes of The Collected Works of John Dewey, 1882-1953 incorporates the tables of contents of all Dewey volumes and includes title and subject indexes. (1728-1) $50.00

Now in paperback...

THE LATER WORKS OF JOHN DEWEY, 1925-1953

Volume 11: 1935-1937
Essays and *Liberalism and Social Action*
Edited by JO ANN BOYDSTON
Introduction by JOHN J. McDERMOTT. (1677-3) $12.95

Volume 12: 1938
Logic: The Theory of Inquiry
Edited by JO ANN BOYDSTON
Introduction by ERNEST NAGEL. (1678-1) $12.95

Add $2.50 when ordering by mail VISA and MasterCard accepted.
Please include 6 25% sales tax on orders shipped to Illinois addresses.

SOUTHERN ILLINOIS UNIVERSITY PRESS
Dept. JP22, P. O. Box 3697, Carbondale, IL 62902-3697

TEMPLE UNIVERSITY PRESS

CHRISTIAN PACIFISM
DANIEL A. DOMBROWSKI
"Well-written. . . A useful statement of an important position." —George Mavrodes, University of Michigan
$34.95

THE CASE AGAINST CHRISTIANITY
MICHAEL MARTIN
$44.95

FAITH, REASON, AND SKEPTICISM
Edited by MARCUS HESTER
Essays by William P. Alston, Robert Audi, Terence Penelhum, and Richard H. Popkin
$39.95

AGAINST INSTINCT
From Biology to Philosophical Psychology
DENNIS M. SENCHUK
$44.95

NOW IN PAPERBACK....

HEIDEGGER AND NAZISM
VICTOR FARIAS
$18.95/$34.95

CAMUS
A Critical Examination
DAVID SPRINTZEN
$18.95/$44.95

THE FORMS OF POWER
From Domination to Transformation
THOMAS E. WARTENBERG
$19.95/$39.95

THE PHILOSOPHY OF ALAIN LOCKE
Harlem Renaissance and Beyond
Edited by LEONARD HARRIS
$17.95/$34.95

THE ROOTS OF THINKING
MAXINE SHEETS-JOHNSTONE
$22.95/$49.95

INTERPRETING WITTGENSTEIN
A Cloud of Philosophy, a Drop of Grammar
RONALD SUTER
$22.95/$44.95

DILEMMAS OF THE AMERICAN SELF
JOHN P. HEWITT
Winner of the 1990 Charles Horton Cooley Award
$18.95/$39.95

TRANSFORMING KNOWLEDGE
ELIZABETH KAMARCK MINNICH
Winner of the 9th Annual Frederic W. Ness Book Award
$18.95/$44.95

THE THEE GENERATION
Reflections on the Coming Revolution
TOM REGAN
$14.95/$24.95

BOOTH 300

order toll-free 1-800-447-1656

NEW FROM THE FREE PRESS

New Third Edition —
Now Available for Immediate Adoption!

GREEK PHILOSOPHY
Thales to Aristotle

Third Edition, Revised and Expanded

Reginald E. Allen, Editor

Widely praised for its concentration on the issues that are most central to the history of Greek philosophy, as well as for its accessibility, *Greek Philosophy: Thales to Aristotle* offers a valuable introduction to the works of the Presocratics, Plato, and Aristotle. Emphasizing metaphysics and epistemology as the foci of Greek philosophy, Allen's book also affords a solid basis for studying the majority of post-classical philosophers who have further explored metaphysics and epistemology as their central concerns.

Including key texts in Socratic ethics, Platonic politics, and natural science, the complete scope of Allen's collection has made it a perennial favorite, the most popular reader available on Greek philosophy. While the third edition retains the earlier emphasis on metaphysics and epistemology, it is enlarged to include more material on Greek science and mathematics, particularly with regards to analysis of infinity. Some of Aristotle's less well-known texts from his Physics have also been added. To complete the picture, revised translations of Plato's Symposium and Aristotle's Categories render these texts more faithful to the original Greek.

Reginald E. Allen is professor of classics and philosophy at Northwestern University.

1991 ISBN: 0-02-900495-0 $14.95 paper

Attention College Professors:
Order your examination copy today! Simply place your request on university stationery, indicating course name, current text, projected enrollment, and the date your next text decision is due
Send to

Fp THE FREE PRESS
A Division of Macmillan/A Maxwell Macmillan Company
Attn Dino Battista,
866 Third Avenue, New York, NY 10022

1992

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 3
MARCH 1992

page
111 *Nominalism and the Contingency of Abstract Objects*
Bob Hale and Crispin Wright
136 *Immaterial Aspects of Thought* James Ross

BOOK REVIEWS
151 Catherine Wilson: *Leibniz's Metaphysics: A Historical and Comparative Study* and
Robert C. Sleigh, Jr.: *Leibniz and Arnauld: A Commentary on their Correspondence* Daniel Garber

166 NOTE TO CONTRIBUTORS
166 ERRATUM NOTICE

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 3, MARCH 1992

NOMINALISM AND THE CONTINGENCY OF ABSTRACT OBJECTS*

TRADITIONAL mathematical nominalism[1] regards a mathematical theory as acceptable only if it can be so interpreted as to involve no purported quantification over or singular reference to abstract objects. To whatever extent a theory fails this test, it is viewed as literally unintelligible. Hartry Field's[2] nominalism is atraditional in this respect. Field disbelieves in abstract objects, of course, and accepts that it is a constraint upon defensible mathematical thought that it avoid commitment to them. But he does not hold it to be a condition on the intelligibility of a mathematical theory that it allow of a non-Platonist *semantics*. On the contrary, he is prepared to grant that a Platonist semantics for, e.g., number theory is descriptively correct: that number-theoretic statements do indeed purport singular reference within and quantification over a specific domain of abstract objects. Since, in his view, there are no such objects, Field has to hold that such statements' truth conditions are systematically unrealized. And so, of course, he does.

How can anyone who holds this view—that math is massively false—hope to save the discipline without semantic reconstruction? By accepting a distinction, roughly parallel to that advocated by Bas van Fraassen[3] in the philosophy of science, between *endorsing* a theory

---

* Thanks to the participants at colloquia at Michigan, Oxford, and the 1991 Frege conference at Munich, organized by Matthias Schirn, at which versions of this material were presented. Special thanks to Hartry Field for much interesting discussion.

[1] See, e.g., W. V. Quine and Nelson Goodman's classic nominalist manifesto, "Steps towards a Constructive Nominalism," *Journal of Symbolic Logic* (1947): 105–22.

[2] As originally expounded in his *Science without Numbers* (New York: Blackwell, 1980).

[3] *The Scientific Image* (New York: Oxford, 1980).

0022-362X/92/8903/111–35

and *accepting it as true*. For van Fraassen, one need not, in endorsing a physical theory, be committed to more than its empirical adequacy—its correctness as far as what is observable is concerned. And for Field, one need not, in endorsing a mathematical theory, be committed to more than its *conservativeness:* where a theory is, roughly, conservative with respect to a discourse just in case any inferences among statements of the discourse which it mediates could at least in principle validly be constructed—though perhaps at great cost in length and notational simplicity—without appeal to that theory.[4] For Field, a mathematical theory can be acceptable, even if its ingredient statements have Platonistic truth conditions, provided it is in this way conservative with respect to *nominalistic* discourse: discourse all of whose ingredient statements have truth conditions that make ontological demands only within the domain of the acceptably concrete. Field's program accordingly has the objective of showing that as much as possible of classical mathematics can be saved as acceptable by this standard.

A question arises, however, about the interpretation of the idea of conservativeness. No mathematical theory that employs a material conditional can be conservative unless consistent. But what understanding is one who disbelieves in abstract objects to have of the notion of consistency? Orthodox proof-theoretic and semantic accounts each involve quantification over nominalistically forbidden entities—*sequences* of sentences in the case of proof-theoretic accounts, and *models* in the case of semantic ones. Field ought, therefore, to forgo both of these approaches. And this is precisely what he does, construing consistency and, associatedly, the notion of consequence utilized in the concept of conservativeness, in *primitively modal* terms. Roughly: a statement is a consequence of others if and only if it is not *possible,* in a primitively modal sense, for it to be false while they are all true, and is inconsistent with others if and only if its negation is a consequence of them.

These proposals have a simple corollary. A consistent theory will be one whose axioms are *possibly collectively true,* where this possibility is to be understood in Field's primitive sense.[5] And a believer in the conservativeness of a particular mathematical theory is committed, on Field's account, to endorsing this possibility.

---

[4] In Field's own formulation: a mathematical theory S is *conservative* if, for any nominalistic assertion A and any body of such assertions N, A is not a consequence of $N + S$ unless A is a consequence of N alone. *Realism, Mathematics and Modality* (New York: Blackwell, 1989), p. 125.

[5] Reflect that a set of sentences—in particular, a mathematical theory—is consistent provided that each of its ingredients, S, is consistent with the remainder, $\{S_1, \ldots, S_n\}$. So if $\{S, S_1, \ldots, S_n\}$ is a consistent theory, then, following through on the foregoing construals, not-S will not be a consequence of $\{S_1, \ldots, S_n\}$, and hence it will be possible for $S_1, \ldots, S_n$ and not-not-S, $= S$, to be simultaneously true.

The objection that is to occupy us arises at this point from the consideration that Field's primitive operator of possibility is *object-linguistic*. It has to be, of course. The program would lose any advantage over more traditional versions of nominalism were it to construe claims about consistency metalinguistically. For if consistency is so construed, a nominalist who believes in the widespread conservativeness, and hence consistency, of mathematics is committed to a belief in nothing other than the feasibility of the traditional nominalist program—of providing concrete (or otherwise nominalistically acceptable) *models* of mathematical theories. And that is just what Field wanted to distance himself from. But when the operator is object-linguistic, Field's belief in the conservativeness of, e.g., number theory would appear to commit him to the view that the Peano axioms, say, might have been true as standardly interpreted, and hence that the nonexistence of numbers is a mere *contingency*. As a nominalist, he disbelieves in the existence of abstract objects. But number theory, he holds, is conservative, so consistent, so—if the reasoning just offered is sound—what its axioms collectively say might have been true. And if that had been so, there would have been numbers.

This may seem odd. But why is it an objection—why should the contingency of numbers and other abstract mathematical objects be awkward for Field? For at least two reasons. First, granted that standard mathematical theories are conservative, and that their deployment in the formulation of, e.g., physical theories can be avoided in the manner Field envisages, there is a question about the possibility of anyone having *grounds* for nominalism in the first place. If whether numbers exist is contingent, the question is presumably beyond a priori resolution. An appraisal will therefore demand empirical evidence. Field agrees with the present authors, however, in rejecting any idea of a *direct*, intuitional or "perceptual" epistemology of abstract objects.[6] And the very conservativeness of number theory, along with its dispensability in formulating scientific theories, would have the effect that there can be no nominalistically statable, *indirect* evidence for or against the existence of numbers. Whether or not the alleged contingency is realized would seem, therefore, to be in principle beyond evidence of any sort, and consequently utterly imponderable. So why is Field a nominalist? Why does he presume to a view on the matter?

---

[6] Along lines, for example, famously suggested—or so it is often supposed—by Gödel (cf. "What is Cantor's Continuum Problem?" in Paul Benacerraf & Hilary Putnam, eds., *Philosophy of Mathematics: Selected Readings* (New York: Cambridge, 1983, 2nd ed.), pp. 483–4); or perhaps as developed by Penelope Maddy, "Perception and Mathematical Intuition," *Philosophical Review*, LXXXIX (1980): 163–96. It would take us to far afield to rehearse our objections to these views here. For some of them, see Wright, "Why Numbers Can Believably Be," *Revue Inernationale de Philosophie*, CLXVII (1988): 425–73, esp. pp. 464–5; and Hale, *Abstract Objects* (New York: Blackwell, 1987).

In his recent book, reacting to this thought, Field writes as follows: "if the dispensability programme *can* be carried out"—that is, if it can be shown not merely that mathematics is conservative but that the hypothesis of mathematical objects is dispensable without loss in explanations, in descriptions of our observations, in accounts of metalogic, and so on:

> that gives us reason to not literally believe mathematics but only to adopt a fictionalist attitude towards it. Admittedly, we can't have *direct evidence* against mathematical entities. We also can't have direct evidence against the hypothesis that there are little green people living inside electrons and that are in principle undiscoverable by human beings; but it seems to me undue epistemological caution to maintain agnosticism rather than flat out disbelief about such an idle hypothesis. I think that platonism has seemed a plausible position because it has been assumed that the existence of mathematical entities is not an idle hypothesis. But if . . . the hypothesis is dispensable without loss . . . it is natural to go beyond agnosticism and assert that mathematical entities do not exist (*Realism, Mathematics and Modality*, pp. 44–5).

One observation about this response would be that the general form of transition it makes, from recognition of the necessary absence of all evidence germane to a hypothesis to a "natural" belief in its negation, is arguably completely unwarranted unless there is independent reason to think that there *would* be germane evidence if the hypothesis were true. That is precisely not the situation in the relevant case. But matters are worse than that. For idle hypotheses, in the sense Field means, have idle negations, too. So "flat out disbelief"—belief in the negation—of what is idle cannot in general be justified merely because it is idle. Agnosticism does not here reflect "undue epistemological caution"; it is demanded by mere consistency.[7]

There is a second and deeper reason, however, why the contingency of abstract mathematical objects might be thought an objectionable outcome. This is the reflection which the contingency in question would seem to be, metaphysically, absolutely *surd*—that there is no prospect of any account of what the alleged contingency could be contingent on. It is deeply rooted in our ordinary conception of contingency that there ought in general to be explanations of *why* things which might have been the case are not the case, and of *why* things are the case which might not have been the case. But in the present kind of instance—the existence of the integers, or real numbers, for example—there is not even a glimmering of how such an explanation might proceed. This seems a strong sign that the notion of contingency would be *misapplied* in the present context, and hence that there is philosophical error in any assumptions that commit Field to so applying it.

---

[7] Or if there is a principle that sanctions disbelief in some such cases, it must involve reference to something more than idleness—and Field does not identify, still less support any such principle.

Field has responded tersely to this criticism, charging that the idea that he suffers any such commitment depends on an equivocation. He writes

> Hale and Wright both note that I regard mathematics, and the existence of mathematical entities, as consistent . . . and that I take consistency as a primitive modal notion, a sort of possibility. They then argue that since I regard it as false that there are mathematical entities, I must hold the existence of such entities to be 'contingently false'; and they both proceed to interpret 'contingently' in some non-logical sense (i.e. some sense other than 'neither logically true nor logically contradictory'), to make the position seem absurd. . . . It should be noted that if [their] objection were good, it would apply equally well against any platonist who is not a logicist: that is, any platonist who agreed that mathematics goes beyond mere logic, and hence that the denial of mathematics is logically consistent and hence 'contingent'. But of course it is not good, for as I've said it turns on an equivocation on the meaning of 'possible' (*Realism, Mathematics and Modality*, pp. 43–5).[8]

Field's response, then, is that the sense of 'possible' for which he acknowledges commitment to the possible truth of the Peano axioms is not one that subserves the objection, which is consequently defused.

But how exactly? There are two—and, so far as we can see, only two—possible lines of thought that would be relevant.

---

[8] The claim that "if [our] objection [i.e., the contingent falsehood objection] were good, it would apply equally against any platonist who is not a logicist" tacitly assumes, in effect, that no relevant notion of necessity (that is, no "absolute" notion) is available, other than the austere notion Field describes. Obviously enough, if it is *not*—save in that austere sense—a contingent matter whether numbers exist or not, or whether or not number-theoretic statements are true (i.e., if they are only *weakly* contingent), then the existence of numbers, and the truth of number-theoretic statements, is, if not impossible (conceptually) necessary. There is thus clear space for a version of Platonism, falling well short of logicism as Field conceives it, that is not prey to the contingent falsehood objection we have urged against Field's own position.

(A) It might be maintained (i) that Field's preferred "austere" notion of logical contingency is the only one available; and (ii) that the objection needs a richer but dubious notion of contingency.

(B) It might be granted that there are available other, less austere but still respectable notions of contingency, but then maintained that while the nonexistence of numbers and other mathematical entities is contingent in the original austere sense, it is not a contingency—a fortiori not a metaphysically surd contingency—in the richer sort of sense demanded by the objection.

The considerations to follow will depend on no assumption about which, if either, of these lines might be Field's own preference.[9] Our interest is rather in the moves available in principle to the nominalism Field has advocated, irrespective of other aspects of its author's actual views. We shall argue that both strategies, (A) and (B), are fraught with difficulty. Since Field's claim—that the original objection is defused by reflecting that it is only to the "logical" contingency of, e.g., numbers that his belief in the conservativeness of number theory commits him—requires that one or the other strategy be defensible, it should be rejected as ill-considered.

The program is as follows. The next section will argue that claims A(i) and A(ii) are highly implausible at best. Section II will briefly explore certain problems with strategy B. This will lead naturally into a discussion, in section III, of certain internal tensions of Field's views about the semantics of mathematical theories with the sorts of consideration by which he seemingly regards Platonism as discredited.

Even if the original objection is not to be dismissed as relying on an equivocation, the nominalist still, of course, has the option of biting the bullet: of allowing that the existence of numbers and other abstract objects is indeed contingent in a richer sense than Field's austerely logical one and of attempting to maintain that the resulting position is unexceptionable. Section IV will attempt to deepen and refine the original objection in various respects. In particular, we shall begin to explore whether the nominalist has any prospect of a satisfactory head-on response to it via play with the idea of *brute contingency*.

I. STRATEGY A

Suppose that there were no other relevant notion of contingency than Field's austerely logical notion—according to which, approximately, a statement is contingent just in case it is neither (first-order) logically true nor (first-order) logically contradictory.[10] How might someone proceed to argue—as per A(ii)—that the contingency objection would then be stifled?

---

[9] Actually, Field has made it quite clear in conversations and correspondence that he would want nothing to do with the former.

One idea might be that, if we introduce 'it is logically contingent that' as a sentential operator, simply and solely as a definitional abbreviation for 'it is neither first-order logically true nor first-order logically contradictory that', we have—so far, anyway—done nothing to provide a sense for statements of the form 'the fact that . . . is contingent upon the fact(s) that . . .', and so we cannot so much as intelligibly formulate the question, "What is the (allegedly) contingent fact (that there are no numbers) contingent *upon*?" Nothing in the way in which we would have introduced the contingency operator would prepare for the idea that when it is (merely) contingent that the things are thus and so, there is—still less has to be—something else upon which their being thus and so *depends*.

This, however, has no force. Asking what the putative contingency, that there are no numbers, is contingent on is a convenient and natural way to raise the kind of question that fuels the objection. But we can just as well ask *why* there are no numbers,[11] given that it is, in Field's austere sense, contingent that there are none. That is a perfectly well-formed and legitimate request for an explanation, for which no further preparation, supplementary to Field's account of contingency, is needed. According to Field, it is logically contingent but false that there are numbers. Well, there are many statements that are logically contingent but false, and it is typically perfectly proper to ask *why* these statements are false. So a special consideration would be needed—and none is so far in sight—to justify rejecting the question as improper in the present case.

---

[10] In fact, this oversimplifies Field's notion. Field's explanation (*Realism, Mathematics and Modality*, pp. 32–8) of "LTrue" (read 'it is logically true that'), in terms of the laws governing it, makes it clear that many statements involving this operator would rank as (austerely) logically true, and that various statements involving second-order quantification would be among them. It is thus strictly incorrect to suggest that a statement is logically contingent for Field if it is neither (first-order) logically true not (first-order) logically contradictory. The fact remains that (austere) logical truth, contingency, and impossibility are, for Field, essentially a matter of logical form—the point is merely that this is not exhaustively characterizable in first-order terms. It should be clear that none of our arguments makes essential play with this simplifying assumption adopted here—in particular, there is no question but that the examples of (austere) logical contingency discussed in this section would rank as such, even given Field's own slightly more generous characterization of the notion.

[11] As indeed, in one of the passages to which Field is responding, Wright in effect does. The passage, which Field actually quotes (cf. *Realism, Mathematics and Modality*, p. 43), runs: "Field has no prospect of an account of what the alleged contingency is contingent *on*. The world does not, in Field's view, but might have contained numbers. But there is no explanation of *why* it contains no numbers; and if it had contained numbers, there would have been no explanation of that either. There are no conditions favorable for the emergence of numbers, and no conditions which prevent their emergence" (Wright, *op. cit.*, p. 465).

Admittedly, there is a gap between its being "typically perfectly proper" to ask why an austerely contingent matter is as it is and there being, on the other, an *internal conceptual connection* ensuring the availability of at least an appropriate *category* of explanation for each austerely contingent matter. We make no claim to have closed this gap, and it is thus theoretically open to a supporter of Field to attempt to exploit it—more about that later. The fact remains that asking for an explanation of why an austerely logical contingency is unrealized is, at least in a wide class of cases, perfectly legitimate. If there are other austere contingencies where the demand for explanation is out of order, there ought to be some stateable distinction between the two sorts of cases. And argument will be needed that the existence of numbers belongs in the latter class. There is no progress otherwise.

It is in any case very implausible that, in line with A(i), a restriction to austerely logical contingency could defensibly be maintained. Any statement is austerely logically contingent whose most explicit first-order formalization is neither a theorem of first-order logic nor refutable therein.[12] So, for instance, both

(i) There are vixens weighing in excess of eighty pounds.

and

(ii) There are male vixens.

rank as austerely logical contingencies. More generally, austerely logical contingency will embrace many statements whose truth it is attractive to regard as, in old-fashioned terminology, *conceptually* impossible but a first-order demonstration of whose impossibility, if one could be given at all, would require supplementary content-explicating postulates of a nonlogical sort. Now, in each of the kinds of example typified by (i) and (ii), respectively, the request for an explanation of *why* the statement in question is false is apparently perfectly intelligible, and it is easy to say how at least the beginnings of an answer should proceed. Given that first-order logic does not preclude the truth of such a statement, the question is, what does? And the answers are, roughly, the typical genetic make-up and feeding habits of vixens, and the nature of the concept of a vixen, respectively.

---

[12] But see note 9.

So much is merely common sense. But the question now is: How should a proponent of A(i) respond to the demand for an explanation of why there are no male vixens? How to respond without betraying an understanding of a broader notion of contingency than the austerely logical variety—a notion of contingency that, if numbers are so contingent, will subserve the original objection, even if, contrary to what was just suggested, austere logical contingency will not?

Well, reflect on the possibilities. In effect, the question—"Why are there no male vixens?"—sets a trilemma:

(i) There is the option of rejecting the demand for an explanation as illegitimate.
(ii) There is the option of attempting an explanation similar to that of why there are no vixens weighing in excess of eighty pounds.
(iii) There is the option of attempting an explanation that preserves a distinction between the two kinds of example.

Option (i) is surely hopeless. It simply is not credible to suppose that there is nothing at all to be said about why there are no male vixens.[13] But it would be equally incredible to suppose that a proper explanation might assimilate the two examples, attributing the non-occurrence of male vixens to *causal* factors, specifically, to the genetic make-up and typical environmental conditions of foxes. Someone who thought that a satisfactory explanation might proceed along those lines would owe a story about what kind of variation in those factors might secure the emergence of male vixens. But there is no such story to tell.

The fact is that only option (iii) is a credible response. More specifically, however indigestible the thought and unpalatable its terminology may be to philosophers weaned on the thin nutrient of W. V. Quine's "Two Dogmas of Empiricism,"[14] the only credible response is to acknowledge that it is not in the nature of the concepts *male* and *vixen* to allow coinstantiation. There may or may not be more to say in response to the request for explanation in this particular case—but that much must be at least the basis of any satisfactory response.

As soon as it is granted, however, that there is work to be done by notions of conceptual possibility and impossibility which apply differentially within the class of austerely logical contingencies, houseroom has been given to just the kind of richer notion of contingency that it was essential to strategy A to repudiate. Say that a statement is *strongly contingent,* just in case it is both austerely logically contingent and not conceptually impossible: no impediment to its truth is generated either by first-order logic or by anything internal to the character of the nonlogical concepts that it involves. If this notion is granted and, with it, the resources for a distinction, among austerely logical contingencies, between those like "There are vixens weighing in excess of eighty pounds" which are also strongly contingent and those like "There are male vixens" which are not, the objection immediately resurges in the form of a dilemma, namely, to which camp should "There are natural numbers" be assigned?

---

[13] A parent who so responded to her child's question would surely be failing to explain something that a responsible parent ought to want to explain.

[14] In *From a Logical Point of View* (Cambridge: Harvard, 1980, rev. ed.).

If to the latter—if the existence of numbers is regarded not as a strong contingency but as a conceptual impossibility—then that is to abandon strategy A in favor of strategy B. But if to the former—if the existence of numbers is conceived as a strong contingency—that immediately invites the assimilation of "There are natural numbers" to "There are vixens weighing in excess of eighty pounds"—a statement whose falsehood allows of a broadly *causal* explanation. And the idea that such an explanation might be possible of why there are no numbers is not merely in tension with our ordinary conception of abstract objects as acausal but is demonstrably inconsistent with the belief that number theory is conservative with respect to nominalistic discourse.[15]

Even if the presumption of *causal* explicability can be defeated, there is no way round the point that to conceive of the existence of numbers as a strong contingency is a commitment to the view that nothing in a correct understanding of number theory or its underlying logic poses any obstacle to its truth. The idea would have to be, rather, that *it is merely a feature of nature* that, if it is, the world is numberless. And to try to take that conception of the matter seriously is simply to point one's chin at the original question: *Why* has nature turned out that way—what kind of consideration might in principle explain the nonexistence of numbers once they are acknowledged to be a strongly contingently nonexisting class of entities? And how can it be legitimate to think of the matter as, so to speak, merely natural, unless that question has some kind of constructive answer?

---

[15] *Proof.* Suppose that a causal explanation of the nonexistence of numbers could be given Then assuming that such an explanation will be of the familiar deductive-nomological pattern, there will be some nominalistically formulated laws (briefly, L) and nominalistic statement of initial or standing conditions (briefly, A) from whose conjunction the statement that there are no numbers could be derived. We may assume that the conjunctive premise is consistent, and that neither conjunct alone entails this conclusion. Let N be elementary number theory Then, since N entails that there are numbers, not$-A$ is a nominalistic statement entailed by $N + L$ but not by L alone, whence—bearing in mind Field's definition (cf. note 4 above)—N is not conservative.

II. STRATEGY B

The question can be avoided by reverting to strategy B. This, recall, involves granting that there is, in addition to any austerely logical sense of contingency like Field's, a richer notion; and then claiming that while the existence or nonexistence of numbers is a contingent matter in the former sense, it is *not* so in the richer sense. The question—"Why are there no numbers?"—would thus admit of a direct answer: the nominalist could simply adduce whatever considerations show that numbers are impossible in the richer sense.

The problem, obviously, is to locate a suitable richer notion—one that allows successful argument that numbers and other abstracta are indeed richly impossible. The simplest line, naturally, would be to attempt to maintain that the existence of numbers is conceptually impossible in the relatively straightforward manner of male vixens, or squares with unequal sides. On this view, all existentially committed classical arithmetical theorems would conceal definitionally elicitable inconsistency. One difficulty with this is that it is crucial to Field's project that arithmetic be formally consistent (else it cannot be conservative). And presumably its formal consistency should be sufficiently robust not to be disrupted merely by expansion of its statements in the light of correct definitions of certain of the concepts involved. But there seems, in any event, no prospect of making the necessary case, i.e., of a successful argument that while there is no overt formal inconsistency in, e.g., the (normally regarded as true) statement that $117 < 5$, it can nevertheless, with the help of acceptable definitions of the specifically arithmetical terms involved, be made to give way to explicitly contradictory consequences, in the manner in which the statement that there are squares with unequal sides gives way to an explicit contradiction on replacement of the geometrical term 'square' by its normal definiens.

If there is to be a conceptual impossibility here, it has to be of a subtler sort. Actually, it is hard to see how the alleged impossibility could be held to derive from anything but the very abstractness of numbers, so that the existence of any kinds of abstract object would have to be impossible. Of course, taken in one way, this would merely inherit the difficulties just glossed. For it seems equally implausible to suppose that the very concept of *abstract object* is self-contradictory after the fashion of *male vixen* or *four-sided triangle*. But perhaps it could be argued that the concept, though not self-contradictory, is nevertheless defective in such a way as to guarantee that it can have no instances.

How? Field's writings endorse *passim* a number of familiar anti-Platonist lines.[16] In particular, he contends—in company with many others, of course—that the acausality and lack of spatiotemporal position of Platonistically conceived mathematical entities make for serious, if not insuperable, difficulties over knowledge of and reference to them. He also endorses the well-known arguments against Platonism given by Paul Benacerraf[17] and by Hilary Putnam.[18] Unless bolstered with some substantial further premise, however, none of these arguments will deliver the type of conclusion that strategy B requires, namely, that there can be no such objects as numbers, sets, etc. For instance, the proper conclusion of a cogent argument based upon some presumed causal constraint on knowledge would be merely that, *if* there are such objects as numbers, we can have no knowledge concerning them. Likewise, the upshot of the arguments —whether based upon some proposed causal theory of reference, or upon the kinds of considerations adduced by Benacerraf or Putnam—purporting to establish that (determinate) reference to numbers is impossible, is just *that,* i.e., that, *if* there are such things as numbers, there is no way we can make (determinate) reference to them. If these arguments are to yield the further conclusion that there simply cannot be any such objects at all, then further premises are needed, perhaps directly to the effect that no objects can exist of which knowledge is in principle impossible or to which there cannot, even in principle, be identifying reference, or perhaps yielding the desired result some other way. In any case, it is clear that proving the impossibility of numbers, or abstract objects generally, is a very tall order which no extant line of philosophical argument meets. If nominalists had such an argument, we would have heard about it.

But if the case were somehow made, there would be a further issue, worth briefly airing. Field, remember, has no quarrel with Platonism as an account of our actual understanding of mathematical language. And this literalism is an attractive and essential element in his kind of nominalism, distinguishing it from nominalist programs of the more traditional reconstructive (and failed) sort. What, in his view, is wrong with Platonism is not that it misrepresents our actual understanding of mathematical statements, or credits them with a metaphysically impossible content, but merely that it tells ontological untruths. But suppose we had the impossibility proof that strategy B demands. What repercussions ought such a discovery, once it were generally known, to have on the semantics of mathematical statements? Could it be appropriate to continue to take the literalist view?

---

[16] *Realism, Mathematics and Modality,* pp. 68–70, also pp. 23 and 230ff.

[17] "What Numbers Could Not Be," *Philosophical Review,* LXXIV (1965): 47–73.

[18] "Models and Reality," *Journal of Symbolic Logic,* XLV (1980). 464–82.

It may seem that exactly this option—combining a literal construal with denying the possible existence of mathematical entities—is what is secured by going *fictionalist*. No doubt it would be absurd to represent a community as *sincerely* making mathematical claims referring to entities that they regarded as impossible. But a face-value construal of their mathematical language need not involve doing so, provided we can avoid interpreting the commitments that their mathematical assertions express as *beliefs in the truth* of those sentences. And that is just what mathematical fictionalism enables us to do. Fictionalism opens a gap between (i) acknowledging that, if the belief prevails in a community that there are, necessarily, no abstract objects, it cannot be part of the best semantic account of the community's mathematical assertions to represent their *makers* as purporting reference to and quantification over such objects, and (ii) abandoning the face-value construal of the terms and quantifiers that their *mathematical language* contains. The gap is opened precisely because fictionalism makes a distinction between endorsing a mathematical statement and accepting it as true.

There is a residual concern, however. Only fictionalism, it seems, can reconcile three crucial ingredients in Field's philosophy of mathematics: (i) a face-value construal of the semantics of number-theoretic language, (ii) the—in his view—justified disbelief that any actual objects answer to the terms and quantifiers discerned therein, and (iii) our endorsement of the axioms and theorems of number theory. But the envisaged case, remember, is one in which we are supposed to be rightly persuaded not merely that there are not but that there *could not* be numbers—the fiction has to be one of acknowledged *impossibilia*. There will therefore be issues about the *intelligibility* of mathematical language, construed at face-value, which fictionalism as such does nothing to address.

The interactions between intelligibility and the perception of various forms of conceptual impossibility are of course subtle. One might believe in the incoherence of the idea of time travel, for instance, yet still, to all intents and purposes, understand and enjoy a fiction in which such technology plays a central part. By contrast, many of us are persuaded of deep conceptual singularities in the idea of the *Cartesian ego* as a temporal continuant—persuaded, e.g., that there is no coherent notion of what it would be for such an entity to endure through a given period of time. Would there nevertheless be no difficulty about devising an intelligible fiction in which such entities were to be the central—repeatedly referred to and quantified over—characters? Again, assuming—plausibly—that there are conceptual obstacles in the way of interpreting an imagined community as genuinely operating with a deviant arithmetic, there is a real question concerning the intelligibility and coherence of a fiction in which the protagonists practice an arithmetic in which $7 + 5 = 13$, etc.[19]

Obviously, in the absence of any clear indication of the kind of conceptual impossibility that a successful prosecution of strategy B might disclose, or the kind of considerations that might be involved, there has to be an element of shadow boxing about this. Nevertheless, there are, unquestionably, kinds of conceptual incoherence that would jeopardize the intelligibility even of fictional reference and quantification. So there can be no a priori assurance that an impossibility "of a more subtle sort" which pursuit of strategy B might disclose will not have such an effect—no assurance that the strategy could be prosecuted without casualty to the semantical literalism that distinguishes Field's position from nominalism of more traditional kinds.

III. LITERALISM AND THE STANDARD ANTI-PLATONIST ARGUMENTS

That there is cause to doubt the coherence of literalism with other aspects of Field's views may be brought out by recalling some of the considerations to which Field actually appeals as discrediting Platonism—in particular, the allegedly irremediable indeterminancy disclosed by Benacerraf's argument and the idea that their acausality and lack of spatial position preclude any satisfactory account of how reference to numbers is possible. For these considerations, if cogent at all, seem by themselves quite sufficient to make trouble for a face-value construal of, e.g., number theory, even in the context of an overall mathematical fictionalism. If refer*ence* to numbers is ruled out on conceptual grounds—whether or not the further step is taken of concluding that the existence of the refer*ents* themselves is impossible—it is immediately cast into doubt whether the face-value construal adds up to a *coherent* semantic story, let alone a preferable one. Nominalism-cum-literalism has it that arithmetical statements, as ordinarily understood, have determinate though unrealized (or only vacuously realized) Platonistic truth conditions. But in order for any statement to have determinate truth conditions, there must be something correct, ergo coherent, to say about the reference of its singular terms, the satisfaction conditions of its predicates, the range of its quantifiers, and so on—and no *coherent* account of these matters, in the nature of the case, can stand in violation of general conditions on the very possession of such semantical characteristics. A face-value construal of mathematical statements—even if offered in a fictionalist spirit—is bound to offend in this respect if it involves attempting to assign reference to the relevant terms in a way that clashes with deep intrinsic features of the reference relation itself; and just that, according to proponents of the Benacerraf and acausality arguments, is what it must involve if it represents those terms as purporting reference to *abstract* objects.

---

[19] Of course, we can readily make sense of a fiction in which the protagonists are depicted as making, more or less systematically, what would, by their own lights, be reckoned arithmetical mistakes; but that is beside the point.

To put the matter another way: one simply has not given a semantics for a discourse if the details of the proposed account controvert essential features of the semantical notions that provide its very framework. But, according to the arguments in question, to represent, e.g., the numerals as having reference, if to anything, then to abstract objects, is precisely so to controvert essential features of the relation of singular reference. Somebody who accepts such arguments should therefore hold that, rather than deliver unrealized (perhaps necessarily unrealizable) truth conditions for the sentences of arithmetic, the face-value construal gives implicit offense against the semantic relations on which it needs to rely in order to say anything determinate about what the truth conditions of arithmetical sentences are—and thus fails to deliver anything at all.

If this is right, then any anti-Platonist driven by the familiar arguments cited approvingly by Field is left, so far as we can see, with just three options. The first is *traditional nominalism*, according to which the appearance of singular reference and quantification presented by mathematical discourse may be genuine, provided the objects referred to and quantified over are concrete. The second is a *reductive* program to disclose that appearance as spurious, and eliminate it. The third is to accommodate the appearance in the only other way that seems open—to go *formalist* and reject altogether the idea that mathematical statements are bona fide contentful statements having genuine truth conditions. On the latter view, such statements possess a determinate syntax, which may be exploited in formal derivations, but no semantics—a fortiori, not a Platonistic one.

What is not apparently an option, for anyone driven by the familiar arguments, is Field's own view, or any view that shares the idea that mathematics, while a fiction, is a *Platonistic* fiction.[20]

IV. BRUTE CONTINGENCIES

Let us see where we have arrived. The literalist can avoid Platonism only at the cost of denying that mathematical theories are true. The question is then unavoidable: Do such theories fail of truth—specifically, do mathematical singular terms fail of reference—necessarily or merely contingently? The preceding suggests that the former option, in effect that pure mathematics is a fiction of impossibilia, is not the way for a supporter of Field to go. It remains to reconsider the alternative, that mathematical and abstract objects generally exist or fail to exist by grace of strong though inexplicable contingency—precisely the position targeted by the main objection.

The objection could be neutralized if it could be shown that the demand for explanation is here illegitimate; that there are, in any case, certain strongly contingent matters—matters to which the case of mathematical objects may properly be compared—of which, however they are resolved, there is no conceivable explanation. "No conceivable explanation" because not every contingency need, of course, be explicable *in fact*. For instance, if only subsumption under ulterior physical laws can explain the obtaining of any particular physical law, then it will be vain to seek an explanation of any *fundamental* physical law, if such there be. Still, there will be *conceptual* space—as it happens unoccupied—for an explanation of such a law; (for the law might not have been fundamental). By contrast, someone who wants to dismiss the question—"Why are there no numbers?"—as inviting an explanation of a matter that is actually a brute contingency, will need to meet the point that there seems to be no corresponding conceptual space—we seem to have no conception of how an answer might run in principle.

Pursuing this direction will involve a two-stage operation: first, a search for other relevant examples of brute contingency—contingencies where we conceive not merely that no explanation need be possible of the direction in which it is resolved but, more, that there simply is no *category* of explanation that might in principle serve; and second, an argument that the existence of numbers, and of abstract objects generally, belongs with this group. What are the prospects?

---

[20] Doubts along the lines of the last few paragraphs are aired by Hale in "Nominalism," in A. D. Irvine, ed., *Physicalism in Mathematics* (Cambridge: Kluwer, 1990), pp. 136–7

In correspondence, Field has remarked:

> On the question of why there are no numbers, my view is that this is like why is there matter, or why are there no immaterial minds, or why is there no God: you simply can't expect there to be an answer to such basic existence and non-existence questions ('basic' in the sense of: applied to a whole 'ontological category').[21]

This is a very natural thought, and it is therefore important to see that it comes to not very much. For unless the claim is to be simply dogmatic, it needs the backing of an *account* of "basic ontological category" which both discloses why explanations of "basic" matters of existence and nonexistence are indeed not to be expected and also ensures that numbers, and abstract objects generally, turn out to be a basic ontological category in the relevant sense. But it is not clear how such an account might run. Indeed, it is obscure what intuitive conception of a basic category covers each of the cited examples. Notice that certain quite attractive proposals will not subserve Field's thought in any case. Persons, and living organisms, will both be basic on many accounts, including—though its author expresses some uncertainty—the largest-class-of-objects-sharing-criteria-of-identity idea of Michael Dummett.[22] But presumably it is not absurd to seek explanations both of why there are any persons and why there is life.

There are, however, other cases which may be held to exhibit brute contingency, and so to provide the model Field needs. Consider the ageless conundrum—"Why is there something, rather than nothing?" An unanswerable question if physical existence is at issue and a satisfactory explanation why a physical state of affairs obtains has to advert to a causally prior situation in which it does not obtain. For a physically empty world would presumably have no causal progeny. Or consider the "great contingency"—the global state of affairs that verifies all actually true contingent propositions.[23] Since the obtaining of each of its ingredients is a matter of contingency, the exact character of the global state is no doubt itself a contingency. But there can be no explaining its exact character—at least not if contingencies can be explained only by reference to further contingencies. For there are, precisely, no *further* contingencies to appeal to; everything further is a matter of necessity.

---

[21] Letter of September 8, 1990.

[22] *Frege: Philosophy of Language* (London: Duckworth, 1973), pp. 75–6.

[23] Here we are indebted to Peter Carruthers and Gideon Rosen for suggestions and helpful discussion.

Do such examples offer any comfort to a supporter of Field? To stress: it will not do just to observe that a certain sort of brute contingency is illustrated by examples of this sort, and is therefore at the call of the nominalist according to convenience. The crux will be whether "There are no numbers," or "There are no abstract objects," comes in the same rather special sort of category. And argument is required. The original objection was that the concept of contingency has no proper application to the cases in question, on the ground that contingencies invite the question—"Why?"—and that we have, in the cases in question, no concept of the form a satisfactory answer might assume. The onus is therefore on the nominalist to show that this is a bad reason—that the correct account of our lack of any such concept is not that the idea of contingency is being stretched beyond the limits of its intelligible application, but that—as in the case of the two mooted examples—explanation is here pre-empted in perfectly intelligible ways, quite compatibly with the claim of contingency.

Reflect on what would be demanded by a serious analogy with those two examples. There the essential work is done by certain restrictive principles: roughly,

> The explanation of the obtaining of a (physical) state of affairs must advert to a causally prior state of affairs in which it does not obtain.

and

> The explanation of a contingency must proceed in terms of further contingencies.

The key point is that such restrictions pre-empt explanation as soon as something sufficiently general is put up as the explanandum, and that this—rather than any misappropriation of the notion of contingency—is at the root of our inability to see how a satisfactory explanation of the contingencies in question might proceed. Accordingly, to establish a strict parallel, the nominalist needs to call attention to a restriction on explanations of mathematical existence which has a corresponding effect: our inability to explain the existence, or nonexistence, of numbers, or abstract objects generally, should be traced to the fact that the demand to do so implicitly transgresses bounds within which a satisfactory explanation would have to work.

The concept of mathematical explanation is a difficult one. But we can at least take it that explaining pure mathematical facts is always a matter of giving proofs, even if not all proofs are explanatory. That might suggest a suitable line for the nominalist. Mathematical proofs, it might be argued, accomplish demonstrations, and therefore explanations, only *within a framework* supplied by the basic assumptions, including assumptions of existence, made by the relevant mathematical theories. Mathematical why-questions, the suggestion continues, are *internal* questions: grant the framework and ontology presupposed by the Peano axioms, and we can explain the infinity of the primes; grant the framework and ontology of Zermelo-Fraenkel set theory and we can—or so it is sometimes thought—explain the truth of the Peano axioms. But just as in the ascent to "Why is there any physical substance at all?" or "Why is the totality of contingencies composed just the way it is?" escalating generality of this sort must sooner or later clash with the possibility of explanation. It has done so by the time we ask why there are (no) mathematical objects.

Although this may seem to pass muster, on closer inspection the analogy limps at the crucial point. What exactly is the restrictive principle in play? Presumably, something like,

> The explanation of any matter of mathematical existence or nonexistence can only be accomplished within a framework in which the existence of a range of mathematical objects is assumed.

But that principle overlooks an asymmetry in the character of mathematical explanations of existence and nonexistence, respectively. No doubt it is only if the existence of the natural numbers is granted that someone can set about trying to explain why the ratio of primes $\leq n$ becomes smaller and smaller as n increases. But what about the explanation offered by Euclid's proof of the *non*existence of a greatest prime? Here it is inessential to the success of the explanation to presuppose that there really are numbers. For the proof succeeds in any case: if the natural numbers exist, then, by Euclid's reasoning, it is in their nature that there is no finite limit to the occurrence of primes; and if they do not, then there are no prime numbers and, a fortiori, no greatest among them.

The point is perfectly general. Mathematical explanations of nonexistence proceed, without existential presupposition, by showing that the corresponding existence claim clashes with essential aspects of our *concept* of the type of mathematical objects concerned, that the concept's being fully and properly exemplified would be inconsistent with the truth of the existence claim. There is no range of objects demanded by all such explanations, whose existence or nonexistence is thereby pushed beyond explanation. The candidate restrictive principle is false.

No firm conclusion can be drawn, of course, on the basis of this one failed comparison. But is there a better one? The heart of the objection is the claim that, while there is no simple connection between contingency and the possibility of explanation, there is nevertheless a *presumptive* connection: explicability is the default status, and the exceptions have to be *explicable* exceptions—something has to be done to show *why* ordinary explanation is precluded. The implicit principle, *P*, is that, for any putative (strong) contingency *C*, it should be possible *either*

(a) to indicate at least a category of explanation that an explanation of *C* might in principle instantiate—to gesture at the lines along which an explanation of *C* might run, even if as a matter of fact no correct explanation of that kind can be given; *or*
(b) to explain why there cannot be any such category—why any explanation of *C* would be required, by virtue of certain special features of the case, to meet unsatisfiable conditions.

And the backing for the principle is merely the natural idea that since, when it is a (strong) contingency that *C*, things might have been otherwise, it ought to be perfectly appropriate to ask *why* they are not otherwise, i.e., why *C*. Since this question is appropriate only if a suitable category of explanation exists, the onus has to be on someone who accepts that no such category exists but who is determined that *C* is nevertheless a (strong) contingency, to show why, in this special case, all space for explanation is closed off, i.e., to satisfy the second disjunct of the principle. This option can be nicely accomplished in the case of the two examples reviewed. But we have so far no clear suggestion how it might be accomplished for the putative contingency that there are no numbers.[24]

All this, however, is undeniably inconclusive. Can we conjure a wider class of examples that may help to place the putative contingency of mathematical objects in a more luminous perspective? We already have

(1) The "great contingency";
(2) "Why is there anything material?"—the Conundrum;
(3) Any fundamental scientific law, *L*;
(4) The existence of God—or, presumably, of any Being whose concept, like that of the Christian God, requires that its existence depend on nothing else.

---

[24] It is important to remember, moreover, that, since we are at this point concerned with a theorist who regards the strong contingency of numbers as the correct account of their modal status, hence as something imposed by an ordinary understanding of number—it has to be possible to draw on *agreed aspects* of that understanding to discharge the obligations of course (b).

Other examples do occur. Suppose physics suggests, for instance, that the universe is approximately *n* years old. Then the question

(5) Why did the universe begin approximately *n* years ago?

will admit of no causal explanatory answer since it presupposes that, prior to that period, there was no physical process—hence nothing to determine that process should start then rather than earlier. Or consider an experiment in which under radioactive decomposition a particle escapes through one slit in a screen rather than another, and suppose that—so we conceive—activity at this micromaterial level is subject only to statistical—indeterministic—law. Then there need be no answer to the question

(6) Why did the particle pass through the top slit?

No doubt such examples put paid to any lingering attraction in the simple thought that, for any bona fide contingency, there has to be something *on* which it is contingent. But collectively they merely reinforce the presumptive connection asserted by the principle, *P*. For in each case we can either indicate quite definite lines along which explanation might have been possible—*L* might have been subsumable under ulterior physical law, for instance, or the laws governing the relevant level of micromaterial activity might have been deterministic—or else, as in cases (1), (2), (4), and (5), we can point to clear conceptual considerations which, drawing just on our ordinary understanding of the example, preclude explanation in principle.

But there *are* more awkward customers. Consider any contingency, *C*, for which, as in cases (3) and (6), we possess a general conception of how an explanation might in principle proceed but where no correct explanation of that sort can actually be given. Then the circumstance—call it *C**—that there is no explanation of *C* will itself be a contingency, and it will be liable to be brute. For how might an explanation of *C** go? Not a priori, presumably, since *C** is a contingency—there might have been an explanation of *C*. But what sort of ***empirical*** explanation might be given of the contingent absence of all empirical explanation of a given (as it happens fundamental) law, or of the (as it happens) indeterministic character of some range of micromaterial events? Thus, for any *C* of the type illustrated by (3) and (6), the corresponding contingency of the form of *C** already poses a problem to the taxonomy implicit in *P*—it will be, prima facie, a case where it is clear neither what a suitable category of explanation might be nor that there can be, a priori, no such category. Thus we can add to the list, for any fundamental law *L*,

(7) There is no explanation of *L*'s obtaining.

and likewise, in respect of example (6),

(8) There is no explanation of the particle's passing through the top slit.

More generally, for any of a wide range of physical phenomena which we can seemingly readily conceive might have been otherwise and yet of which our best current conception is that they have no physical preconditions—the whole phenomenon of gravitational attraction, for instance, or space's being non-Euclidean, or *n*-dimensional; in all such cases, the claim that there is no explanation of the phenomenon will seemingly be itself a contingency about which we are clear neither along what general lines an explanation might be given nor about what if any conceptual considerations preclude explanation in principle.

Well, are not these reflections enough to dislodge principle *P*? Do we not have to admit now that the intuition of contingency can perfectly justifiably be robust in circumstances where we are unclear whether either limb of principle *P* can be satisfactorily made out? Of course, these, as it were, second-order contingencies are of a special sort, and bear no immediate analogy to the putative contingency that there are no abstract objects. Still, the principle that drives the objection is not, it appears, exceptionless in any case. So what cause for skepticism if the nominalist postulates a further class of exceptions?

In an earlier discussion Wright wrote:

> . . . the idea of a contingent state of affairs that is contingent on nothing, admits of no evidence even in principle, and is beyond anything we could justifiably regard as cognition, is simply not a credible piece of metaphysics. . . .[25]

Focusing, as we have been, on the question of explanation belongs with the thought that bona fide contingencies are presumptively contingent on something, and that the exceptions should be few and clearly explicable as such. But we should not lose sight of the second component in the quoted complaint—the utter epistemic isolation of the existence or nonexistence of mathematical objects, when construed as a (strong) contingency. Moreover, it is important to be clear that the offense given thereby need not depend on any implicit verificationism—any general aversion to the idea of evidence-transcendent fact. The reason there can be no cognition of, nor evidence for or against, the existence of mathematical objects, conceived as a strong contingency, is because *nothing whatever*—in particular, no uncontroversial contingency—varies as a function of their existence or nonexistence, save precisely that.[26] It is merely one consequence of this that their existence or nonexistence exerts no differential influence on what is cognitively possible. The conception that mathematical objects exist or not as a matter of strong contingency thus clashes not merely with the presumption that contingencies should be contingent on something but, equally, with the converse principle that bona fide contingencies are presumptively *things on which other bona fide contingencies are contingent.*

---

[25] Wright, "Why Numbers Can Believably Be," p. 465; cf. also Hale, *Abstract Objects*, pp. 111–2.

There is, of course, a vast range of ordinary unproblematic contingencies that respect both these conditions. That is the normal case. But what is striking is that *all* the examples we have considered, even those which prima facie make trouble for *P*, respect the converse principle. There are manifold uncontroversial contingencies that would not be so but for the detail of the "great contingency," the existence of matter, the obtaining of a particular scientific law, the universe's being *n* years old, space's being *n*-dimensional, the particle's going through the top slit (it leaves traces), God's existing (if he does), *L*'s being fundamental (for if *L* were not fundamental, other superior laws would obtain and would be manifest in differential behavior of matter), and there being no explanation why the particle went through the top slit (for if there were an explanation, the relevant class of phenomena would not be indeterministic and all sorts of things would be different).

The suggestion, then, is that both the following presumptive principles constrain the ascription of strong contingency:

(1) That a putative contingency will be contingent on something—there will be circumstances on which its obtaining depends (and whose citation can thus contribute to its explanation).
(2) That a putative contingency will have things contingent on it—there will be circumstances whose obtaining depends on it (would be otherwise if it did not obtain.)

Brute contingencies are counterexamples to the first. Call a contingency *barren*, by contrast, if it is a counterexample to the second. Then none of the brute contingencies reviewed in our list of eight seems to be barren. Uncontroversial examples of barren contingencies are, indeed, hard to muster. The occurrence of a mental state or episode, viewed in the special way characteristic of epiphenomenalism, would be one possible type of example. But even that would not also, of course, be brute. What is striking about "There are (no) numbers," viewed as a strong contingency, is that, at least with respect to all other uncontroversially contingent matters, the contingency it putatively expresses is absolutely *insular*—is both barren *and* brute. We conjecture that there is no other example of a contingency—at least, no case that would pass without controversy as a contingency—that has both those features.

---

[26] If this were not so, then mathematical theories would not be conservative with respect to nominalistically stateable consequences.

It would be satisfying to disclose some definite metaphysical incoherence in the notion that there can be absolutely insular contingencies. At the very least, it is at odds with the idea that the realm of contingency forms a single integrated system—a tree-like structure in which the connecting links are dependencies and in which every node is linked to others. There is a potential place in such a structure for initial nodes—nodes ancestrally dependent on no others, which thus correspond to brute contingencies—and for terminating branches, whose terminal nodes thus correspond to barren contingencies. What there is no place for is a node that is both initial and terminal; for by forgoing any relations of dependency, such a node forfeits any case to be regarded as part of the structure at all.

What should we conclude? The picture of the realm of contingency as comprising such a unified web seems to go deep enough in our ordinary thinking to ensure that it will be very difficult to vindicate the conception of mathematical objects as strongly contingent merely by description of anything we already find intuitive. But it is hard to avoid feeling that there is a deeper incongruity in any play with absolutely insular contingency, yet to be brought out. At the very least, a supporter of Field who goes in this direction has to swallow some very lumpy metaphysics for the sake of his ontological conscience.

The main purpose of this paper has been to raise the question whether someone who takes a nominalist-cum-literalist position has the resources coherently to classify the modal status of the proposition that there are numbers, or abstract mathematical objects generally. By contrast, the form of deflationary Platonism—Fregean Platonism, or "Platonism for cheap," as Field styles it—whose claims we have each defended elsewhere, does, we believe, provide a coherent and coherently argued view: it is that the existence of numbers is a necessity, whose truth may be recognized a priori. But the further defense of that contention is for another occasion: in particular, we ought to stress that the argument of this paper is not, and is not meant to be, a positive argument for this, or any other, version of Platonism.[27] On the contrary, it relies upon no assumption that could not be endorsed by an adherent of classical reductive nominalism, formalism, or—so far as we can see—virtually any other position that stands opposed to the unorthodox variety of nominalism against which it is directed.

CRISPIN WRIGHT

University of St. Andrews and Michigan

BOB HALE

University of St. Andrews

---

[27] Experience has taught us that there is a tendency toward misunderstanding of this point. So it may help to stress what sort of view comes within the range of our argument. Not any form of anti-Platonism but only views which specifically allow four things:

(i) That the sentences of number theory, analysis, etc., have sufficiently determinate truth conditions to give sense to the question whether they are satisfied or not;
(ii) that those truth conditions are correctly reflected by the overt syntax of the sentences in question;
(iii) that those truth conditions are systematically unrealized; and
(iv) that they are so as a matter either of conceptual necessity or strong contingency.

—only an anti-Platonism that grants *all* these is in trouble if the argument of the paper fully succeeds. Our discussion, if cogent, ought therefore to leave most anti-Platonists undismayed, just because the effect of the most widely influential anti-Platonist arguments—in particular, those deriving from causal accounts of singular reference and Benacerrafian worries about the inscrutability of reference of mathematical singular terms—is (or ought to be) to call one or more of these presuppositions into question. Benacerraf himself, for instance, rejected clause (ii); and section III argued in effect that one who holds that reference to abstract objects is necessarily precluded by their causal inertness cannot coherently endorse both (i) and (ii).

IMMATERIAL ASPECTS OF THOUGHT

ANIMAL cognition and desire, from the appetite of a clam to the optical systems of vultures and frigate birds, is supposed to have neurobiological explanations resultant from, if not reducible to, universal laws of physics. That is a minimal and modest project for epistemology naturalized, one to be assisted by specialized sciences.[1]

There is a larger and bolder project of epistemology naturalized, namely, to explain human thought in terms available to physical science, particularly the aspects of thought that carry truth values, and have formal features, like validity or mathematical form. That project seems to have hit a stone wall, a difficulty so grave that philosophers dismiss the underlying argument, or adopt a cavalier certainty that our judgments only simulate certain pure forms and never are real cases of, e.g., conjunction, ***modus ponens***, adding, or genuine validity. The difficulty is that, in principle, such truth-carrying thoughts[2] cannot be wholly physical (though they might have a physical medium),[3] because they have features that no physical thing or process can have at all.[4]

---

[1] After three centuries of amazingly successful science, we do not have a successful explanation of animal cognition, not even for a spider or a fish. Probably, we have been misconceiving the project in ways that makes science both less productive and less helpful.

[2] Thinking here means "judgmental understanding"—what Aristotle thought to be the actuality of the intellect (*De Anima*, bk. III, ch. 4, 429b, 30: "Mind is in a sense potentially whatever is thinkable, though actually it is nothing until it has thought"). There are many kinds of thinking; some thinkings are bodily doings, like my pouring a liquid. But it is only the processes of understanding that I am now trying to show cannot be wholly physical; understandings that involve feeling cannot be entirely nonphysical either, any more than my going for a walk can be a mere willing.

[3] See Aristotle's argument (*De Anima*, bk. III, ch. 4, 429a, 10–28; see also Aquinas's commentary in *Aristotle's De Anima in the Version of William of Moerbeke and the Commentary of St. Thomas Aquinas*, Kenelm Foster and Silvester Humphries, trans. (New Haven: Yale, 1959 repr.), sec. 684–6, pp. 406–7) that the understanding cannot have an organ as sight has the eye (and nowadays philosophers suppose thinking has the brain), because the limited physical states of an organ would fall short of the contrasting states of understanding that we know we can attain.

[4] Philosophers should not recoil with distaste at such remarks about thought, because they attribute even odder features to propositions, e.g., being infinite in number, belonging to a tight logical network with formal features like "excluded middle," and being such that every one is determinately either logically related, by implication or exclusion, or logically independent of every other; in fact, in a system of material implication, no proposition is logically independent of any other.

0022-362X/92/8903/136–50

I propose to articulate that "difficulty in principle" so as to press home the point that it cannot be dismissed or evaded, or the underlying arguments or costs disregarded. First, the underlying arguments themselves are among the jewels of analytic philosophy (underdetermination considerations); and, secondly, to deny that our judgments are of definite logical forms and pure functions conflicts with our own certainty and with what we tell our logic, mathematics, and linguistics students about validity, proof, and formal syntax, and leaves us unable to explain what we do when we do mathematics, logic, or any other formal thinking.

But now let us look at the argument:

> Some thinking (judgment) is determinate in a way no physical process can be. Consequently, such thinking cannot be (wholly[5]) a physical process. If all thinking, all judgment, is determinate in that way, no physical process can be (the whole of) any judgment at all. Furthermore, "functions" *among* physical states cannot be determinate enough to be such judgments, either. Hence some judgments can be neither wholly physical processes nor wholly functions among physical processes.

Certain thinking, in a single case, is of a definite abstract form (e.g., $N \times N = N^2$), and not indeterminate among incompossible forms (see I below). No physical process can be that definite in its form in a single case. Adding cases even to infinity, unless they are all the possible cases, will not exclude incompossible forms. But supplying all possible cases of any pure function is impossible. So, no physical process can exclude incompossible functions from being equally well (or badly) satisfied (see II below). Thus, no physical process can *be* a case of such thinking. The same holds for functions among physical states (see IV below).

I. THE DETERMINATENESS OF SOME THOUGHT PROCESSES

Can judgments really be of such definite "pure" forms? They have to be; otherwise, they will fail to have the features we attribute to them and upon which the truth of certain judgments about validity, inconsistency, and truth depend; for instance, they have to exclude incompossible forms or they would lack the very features we take to be definitive of their sorts: e.g., conjunction, disjunction, syllogistic, *modus ponens*, etc. The single case of thinking has to be of an abstract "form" (a "pure" function) that is not indeterminate among incompossible ones. For instance, if I square a number—not just happen in the course of adding to write down a sum that is the square, but if I actually square the number—I think in the form "$N \times N = N^2$."

---

[5] But in part, yes, in the sense that my utterances are physical. Moreover, the thought may not even be possible apart from feeling or sense, just as a gesture is not possible without bodily movement. The target in this paper is theories that thoughts are "no more than" physical or functions determined physically; not that, for us, they are "at least physically realized."

The same point again. I can reason in the form, *modus ponens* ("If *p* then *q*"; "*p*"; "therefore, q"). Reasoning by *modus ponens* requires that no incompossible form also be "realized" (in the same sense) by what I have done. Reasoning in that form is thinking in a way that is truth-preserving for all cases that realize the form. What is done cannot, therefore, be indeterminate among structures, some of which are not truth preserving.[6] That is why valid reasoning cannot be only an approximation of the form, but must be of the form. Otherwise, it will as much fail to be truth-preserving for all relevant cases as it succeeds; and thus the whole point of validity will be lost. Thus, we already know that the evasion, "We do not really conjoin, add, or do *modus ponens* but only simulate them," cannot be correct. Still, I shall consider it fully below.

"Being truth preserving for *all* relevant cases" is a feature of the single case. The form of the reasoning that *actually* occurs is "truth-preserving," regardless of which case it is. Otherwise, it would not be "impossible by virtue of the form to proceed from truth to falsity" in *that* reasoning (especially when the premises are not true). Thus, the form of the actual "encompasses" (logically contains) all relevant counterfactual situations. In fact, it encompasses all relevant cases whatever. Without that, there is no genuine difference between valid and invalid reasoning.

Squaring, conjoining, adding. I propose with some simple cases to reinforce the, perhaps already obvious, point that the pure function has to be wholly realized in the single case, and cannot consist in the array of "inputs and outputs" for a certain kind of thinking. Does anyone doubt that we can actually square numbers? "4 times 4 is sixteen"; a definite form ($N \times N = N^2$) is "squaring" for all relevant cases, whether or not we are able to process the digits, or talk long enough to give the answer. To be squaring, I have to be doing something which works for all the cases, something for which any relevant case can be substituted without change in what I am doing, but only in which thing is done.

Size and length of computation, for example, are external to the form of thinking, accidental to what is done. I am squaring just in case my thinking is of the form mentioned. If it is of any incompossible form, or is indeterminate among incompossible forms, it is *not* of the form, "N times $N = N$ squared." It is not then squaring, however much its products may look like it, and however long a sequences of its outputs do.

---

[6] I am not, of course, suggesting that a valid course of reasoning is not also a case of a variety of invalid forms, e.g., "*P*, therefore, *C*." But it must determinately be a case of some valid form.

The fact that I cannot process every case of *modus ponens,* because most of them have premises too long for me to remember, sentences too long to say, or words I do not understand, is adventitious, like my not being able to do *modus ponens* in Portuguese. Those are features of the functors, not of the function. The function that has to be realized in every case is the one wholly realized in the single case.

That point is to be taken literally: that the function is wholly present, not by approximation, exemplification, or simulation, but by realization in the single case. To make that distinction clearer, consider an even simpler function, "conjoining." Conjoining is the functional arrangement of an n-tuple of assertions into a single assertion that is determinately true just in case every one of the n-tuple of judgments is, and false otherwise. The truth of the whole block is the truth of all of the units ("$p \cdot q = T$ just in case $p = T$ and $q = T$"). I can conjoin every sentence in the fourteenth edition of the Encyclopedia Britannica, or yesterday's *Times*. What I do in the single case is what would conjoin any string of suitable units, even ones too long for me to think of, or beyond my access to refer to. It is impossible to *conjoin* thoughts, if what I do is indeterminate among incompossible forms (at the same level).

Adding—genuinely adding, not estimating—is a sum-giving thought form for any suitable array of numbers.[7] If I add two "elevens," I am doing what would have given "forty-four" had I been adding two "twenty-twos" (and not making mistakes), and so on for every other combination of suitable numbers. I cannot be really adding when I do something which gives the "right output" but which cannot, by its form, determine the "right outcome" for any case whatever, even one on which I make a mistake. There is a great difference between adding incorrectly and doing something else, like guessing, estimating, or following a routine or algorithm.

---

[7] Some conjunction tasks seem possible that are not: e.g., to conjoin all statements that can be expressed in English. That impossibility is not because of some fuzziness about the function "conjoin," but because the supposed totality is incoherent. You cannot add up all the even numbers, taken pairwise, just as you cannot conjoin all the sentences of English. See note 10.

The adding I am talking about, like conjoining, is a form of understanding.

This is not a claim about how many states we can be in. This is a claim about the ability exercised in a single case, the ability to think in a form that is sum-giving for every sum, a definite thought form distinct from every other. When a person has acquired such an ability is not always transparent from successful answers, and it can be exhibited even by mistakes.

Definite forms of thought are dispositive for every relevant case actual, potential, and counterfactual. Yet the "function" does not consist in the array of inputs and outcomes.[8] The function is the form by which inputs yield outputs. The array of inputs and outputs for a function is the logical tail of the comet, not what the function is.[9]

The *trait* that determines the tail of the comet, the trait that "settles every relevant case, including all countercases," marks the contrast with any physical process: a physical process has no feature that can do that. That grounds my main argument: that a necessary consequence of even a single case of such thinking is something that is logically impossible to be a consequence of *any* physical process, or function among physical processes, whatever. Thus, the activity of such thinking cannot be a physical process, and the ability for such thinking cannot be a physical capacity.

II THE INDETERMINACY OF THE PHYSICAL

Now we need reasons why no physical process or function among physical processes can determine "the outcome" for every relevant case of a "pure" function. Those considerations mark some of the most successful analytic philosophy, from W. V. Quine, to Nelson Goodman, to Saul Kripke. No physical process is so definite as to determine among incompossible abstract functions that one rather than another is realized, and thus to settle for every relevant case what the "outcome" is to be. That indeterminacy remains no matter how long the physical process is "repeated," even infinitely. In a word, with a machine it is indeterminate among incompossible functions what it is doing, no matter what it does.[10] Therefore, no matter what it does, what it is doing remains formally indeterminate. Goodman's[11] "grue" considerations and the plus-quus adaptations by Kripke[12] suggest the form of my argument to show that. The argument is as follows.

---

[8] We can even add certain nonterminating decimals, like 33333 and .66666 carrying from infinity to get 1. That is a form of understanding.

[9] Equivalent but nonsynonymous functions would give the same arrays from inputs to outputs. Besides, a device that went to an address for the answer, and took it out in an envelope (encoded), which it did not open (decode) but handed to you (displayed for you to decode), could be made to produce the same array of outputs as addition Yet it would not be adding. Besides, look at this function: 10 Z = X*X*X; 20 Print Z; 30 X = X + 1; 40 GOTO 10. That is a machine function for an endless loop to print the cube of every number beginning with zero. You can *see* that no matter what outputs the machine gives, it might have been doing something other than printing successive cubes, unless it produces *all* cubes—which cannot be done.

Whatever the discriminable features of a physical process may be, there will always be a pair of ***incompatible*** predicates, each as empirically adequate as the other, to name a function the exhibited data or process "satisfies." That condition holds for any finite actual "outputs," no matter how many. That is a feature of physical process itself, of change. There is nothing about a physical process, or any repetitions of it, to block it from being a case of incompossible forms ("functions"), if it could be a case of any pure form at all. That is because the differentiating point, the point where the behavioral outputs diverge to manifest different functions, can lie beyond the actual, even if the actual should be infinite; e.g., it could lie in what the thing would have done, had things been otherwise in certain ways. For instance, if the function is $x(*)y = (x + y$, if $y < 10^{40\ \text{years}}$, $= x + y + 1$, otherwise), the differentiating output would lie beyond the conjectured life of the universe.

Just as rectangular doors can approximate Euclidean rectangularity, so physical change can simulate pure functions but cannot realize them. For instance, there are no physical features by which an adding machine, whether it is an old mechanical "gear" machine or a hand calculator or a full computer, can exclude its satisfying a function incompatible with addition, say, quaddition (cf. Kripke's definition (*op. cit.*, p. 9) of the function to show the indeterminacy of the single case: quus, symbolized by the plus sign in a circle, "is defined by: $x \oplus y = x + y$, if $x, y < 57$, $=5$ otherwise") modified so that the differentiating outputs (not what constitutes the difference, but what manifests it) lie beyond the lifetime of the machine. The consequence is that a physical process is *really* indeterminate among incompatible abstract functions.

---

[10] Postulating an infinity of cases will not suitably discriminate the functions that are the same for even numbers but differ for odd numbers after N. Postulating that "all" the cases are actual involves an ***incoherent totality,*** because the machine cannot both do all that it does ***and*** all that it might have done instead. Consequently, a pure function does not reduce to a pattern of inputs and outputs.

"All the additions" is as incoherent as "all the sets." So "what" addition is cannot be explained by "all the outcomes": rather, each and every outcome is determined by what addition is. It is impossible that all cases of addition be actual, even if infinities are performed because, even if we used up all the suitable numbers, the function itself would still be repeatable, say, for the same additions, but now done in a different order. The function cannot be exhausted by its cases, however many there are.

[11] See "The New Riddle of Induction," in ***Fact, Fiction and Forecast,*** 2nd ed. (Indianapolis: Bobbs-Merrill, 1968), pp. 63–86.

[12] ***Wittgenstein on Rules and Private Language*** (Cambridge: Harvard, 1982), p. 9. and ***passim.***

Extending the list of outputs will not select among incompatible functions whose differentiating "point" lies beyond the lifetime (or performance time) of the machine. That, of course, is not the basis for the indeterminacy; it is just a grue-like illustration. Adding is not a sequence of outputs; it is summing; whereas if the process were quadding, all its outputs would be quadditions, whether or not they differed in quantity from additions (before a differentiating point shows up to make the outputs diverge from sums).

For any outputs to be sums, the machine has to add. But the indeterminacy among incompossible functions is to be found in each single case, and therefore in every case. Thus, the machine *never* adds.

Extending the outputs, even to infinity, is unavailing. If the machine is not really adding in the single case, no matter how many actual outputs seem "right," say, for all even numbers taken pairwise (see the qualifying comments in notes 7 and 10 about incoherent totalities), had all relevant cases been included, there would have been nonsums. Kripke drew a skeptical conclusion from such facts, that it is indeterminate which function the machine satisfies, and thus "there is no fact of the matter" as to whether it adds or not. He ought to conclude, instead, that it is *not* adding; that if it is indeterminate (physically and logically, not just epistemically) which function is realized among incompossible functions, *none* of them is. That follows from the logical requirement, for each such function, that any realization of it must be of it and not of an incompossible one.

There is no doubt, then, as to what the machine is doing. It adds, calculates, recalls, etc., by simulation. What it does gets the name of what we do, because it reliably gets the results we do (perhaps even more reliably than we do) when we add by a distinct process. The machine adds the way puppets walk. The names are analogous. The machine attains enough reliability, stability, and economy of output to achieve realism without reality. A flight simulator has enough realism for flight training; you are really trained, but you were not really flying.

A decisive reason why a physical process cannot be determinate among incompossible abstract functions is "amplified grueness": a physical process, however short or long, however few or many outputs, is compatible with counterfactually opposed predicates; even the entire cosmos is. Since such predicates can name functions from "input to output" for every change, any physical process is indeterminate among opposed functions. This is like the projection of a curve from a finite sample of points: any choice has an incompatible competitor.

We have no doubt that the processes in a mechanical adding machine and in a personal computer are entirely physical. Addition cannot be identical with either of those physical processes because then it could not be done by the other. Suppose that addition is identical with a function among those processes. Then the processes would have to determine that function to the exclusion of every incompossible function. But they cannot do that, as the "quus," "grue," and "points-on-a-curve" examples show. So the machines cannot really add.

Secondly, opposed functions that are infinite (that is, are a "conversion" of an infinity of inputs into an infinity of outputs) can have finite sequences, as large as you like, of coincident outputs; they can even have subsequences that are infinitely long and not different (e.g., functions that operate "the same" on even numbers but differently on odd numbers). So for a machine process to be fully determinate, every output for a function would have to occur. For an infinite function, that is impossible. The machine cannot physically do everything it actually does and also do everything it might have done.[13] That is the heart of the matter. The physical, as pro-

---

[13] There is a complementary line of inquiry about immateriality. Christopher Cherniak argues (*Minimal Rationality* (Cambridge: MIT, 1986), p. 127) that because a physical object cannot be in an infinity of states, the mind treated as a brain computer is of limited understanding. That would be an understatement, were it true. Most of what actually happens would be unintelligible to us. An infinity of English sentences would be unintelligible, as would "most" truths of arithmetic.

For even if each of the finite number of electrochemical states the brain is capable of realizing actually happened, say, 10^{140} different thoughts, there would be an infinity of mathematical theorems we could not even *understand* because there would be no brain state or function among brain states to realize them.

The opposite seems to be true: there is nothing that is in principle unintelligible insofar as it has being, as Plato and Aristotle both thought. And we are able to be in an infinity of states of understanding, not successively but qualitatively. That is, we have the active ability to understand *anything* (accidents of presentation and of intelligence quotient being ignored for now). Thus, there is no arithmetical theorem we cannot understand, accidents ignored for now. Nor are there any cess, is formally vague, no matter how far you extend it, or how minutely you describe its innermost mechanisms. The conclusion is that a physical process cannot realize an abstract function. It can at most simulate it.

What Happened to Nature? Do natural processes, say, the behavior of a freely falling body, not realize pure functions like "$d = 1/2gt^2$" and, where $g = 32$, "$d = 16t^2$"? And is it not true that an object in empty space decreases in length in the direction it is traveling by an amount equal to $\sqrt{1 - v^2/c^2}$ There are two reasons why such processes do not realize pure abstract functions of the sorts mentioned, only the second being relevant to the present discussion. First, these laws apply by idealization. What is "*the* direction" in which the object is traveling? There are no "point masses." That is an idealization, as is its rest mass (say, for photons or neutrinos, which are always moving at *C*). No object falling to earth is in a vacuum and under no gravitational attraction to other bodies. Physical phenomena often come close to our mathematizations which, of course, are invented to represent them. But those mathematizations are idealizations.[14] That the laws are idealizations does not affect the present point.

The kind of indeterminacy I am talking about is different from that. For the incompossible functions are equally idealizations, and may differ only logically because the "manifestation phenomena" lie

---

well-formed utterances of any of the conjectured 10,000 human languages (most now lost) that we would not understand in the appropriate circumstances. But any one of those languages would require more than all the brain states. Brain states would have to be vehicles for varying content, perhaps media for thought and not the *same* thing.

Nothing is excluded because of its subject matter. Ours is not a successive infinite capacity (if we do not exist forever) but a selective infinite capacity. That is why the brain cannot even be the organ of thought, the way the eye is the organ of sight, as Aristotle, Avicenna, Averroes, Aquinas, and many others argued; otherwise, there would be something (that might be actually) that is unintelligible. Our corporeality imposes accidental limitations on understanding, the most important of which is that our contents of judgment have to be made by dematerialization (abstraction) and our intelligence cannot directly access immaterial being (e.g., angels or God). One consequence is the indeterminacy of contingent truth (see note 17).

How the dematerialization involved in our understanding something as shape (without consideration of which thing it is, or of its particular material composition) or our understanding something as in being (without consideration of its being material) could even come about is totally beyond the resources of any known experimental or formal science.

[14] See Nancy Cartwright, *How the Laws of Physics Lie* (New York: Oxford, 1983), and Ian Hacking, *Representing and Intervening* (New York: Cambridge, 1983).

beyond the actual (it being presupposed that all the actual phenomena accord with each function). So it is not a consequence of this account that there are no general and mathematizable laws of nature. Rather, just because there are general and mathematizable regularities, an object falling to the earth "in a vacuum" satisfies some incompossible function just as well as it satisfies "$d = 1/2gt$." That is a consequence of the underdetermination arguments.

Now, to accept the overall argument, one does not need to deny that there are definite natural structures, like benzene rings, carbon crystals, or the structural (and behavior-explaining) molecular differences among procaine, novocaine, and cocaine. These are real structures realized in many things, but their descriptions include the *sort* of matter (atoms or molecules) as well as the "dynamic arrangement." They are not pure functions.[15]

A musical score, say, Mahler's 2nd Symphony, can be regarded as an analog computer that determines, from any given initial sound, the successive relative sounds and their relative lengths (within conventions of intervals and length), and thus is a function from initial sound onto successor sounds; yet, from the sounds (a performance) there is not a unique score determined among incompossible ones, except by convention. So, too, when we abstract the formal structure, without matter, the physical thing (cell, molecule, gene, enzyme) or process will satisfy a logically incompossible structure just as well.

III. RETREAT FROM PEOPLE

So, to avoid the argument, someone will say:

> We do not really add, either; we just simulate addition. Pure addition is just as much an idealization as $E = mc^2$. Of course, we can define such pure functions but cannot realize them; that is just a case of the many functions we can define which cannot be computed by any finite automation, or any other computer either. In a word, the fact that there is no pure addition and no pure conjunction or *modus ponens* is no odder than the fact that there are no perfect triangles.

We cannot really add, conjoin, or do *modus ponens*? Now that is expensive. In fact, the cost of saying we only simulate the pure functions is astronomical. For in order to maintain that the processes are basically material, the philosopher has to deny outright that we do the very things we had claimed all along that we do. Yet our doing these things is essential to the reliability of our reasoning. Moreover, we certainly can, Platonistically, define the ideal functions, otherwise we cannot say definitely *what* we cannot do. That exposes a contradiction in the denial that we can think in pure functions, however; for to define such a function is to think in a form that is not indeterminate among incompossible forms. To become convinced that I can only simulate the recognition that two Euclidean right triangles with equal sides are congruent, I have to judge negatively with all the determinateness that has just been denied. Each Platonistic definition of one of the processes, and each description of the content of logical or arithmetical judgment, is as definite a form of thought as any of the processes being denied; and each judgment that we do not do such and such a function is as definite in form as is conjunction, addition, or any of the judgments that are challenged; otherwise, what is denied would be indeterminate. It is implausible enough to say we do not really add or conjoin. It is beyond credibility to say we cannot definitely *deny* that we add, conjoin, assert the congruence of triangles, or define particular functions, like conjunction.

---

[15] General natures (e.g., structural steel) do "have" abstract forms, but are not "pure functions." Two humans, proteins, or cells are the same, not by realizing the same abstract form, but by a structure "solid" with each individual (but not satisfactorily described without resort to atomic components) that does not differ, as to structure or components, from other individuals. There can be mathematical abstractions of those structures, many of which we can already formulate (cf. *Scientific Tables* (Basel: CIBA-GEIGY, 1970)).

The final and greatest cost of insisting that our judgments are not more determinate as to pure functions than physical processes can be, is that we can do nothing logical at all, and no pure mathematics either. Now, who believes that?

There is not some parallel evidential indeterminacy between our activities and those of a machine whereby we cannot be sure what either is doing.[16] The machine cannot in principle add. We can be sure of that. And we can, and do add, and conjoin and reason syllogistically. We can be sure of that, too.

Someone rejoins, "So you say. But we might be just simulating." The rejoinder defeats itself. By its presumption, it grants the force of the argument as a whole, that there are pure functions and that, if certain thought processes were physical processes or functions among them, they would not be formally determinate. It merely asserts as a counterpossibility that I may think I am adding, etc., when I am only simulating a pure function. But to think I am adding or conjoining, with a clear idea of *what* that is, is to perform a pure function in that very thought, whether it is true or not.

---

[16] I think Kripke (*op. cit.*, pp. 21, 65, and 71) interprets what he regards as indeterminacy as to whether I meant plus or quus as the basis for alleging an indeterminacy about what I do. ("There is no fact of the matter.") I say this gets the explanatory order backward and invites mistaken conclusions.

Besides, such counterpossibilities require an ontological status for the pure functions simulated. We think of them and even define them. If that is so, then the thoughts and definitions cannot be indeterminate among incompatible functions because no definite function would then be defined by such thinking. So those function-determining thoughts cannot themselves be just simulations but have to realize pure functions, e.g., "defining addition," "conceiving *modus ponens.*" Hence, in order to be mistaken in a certain way, I have to think in exactly the way that cannot be entirely physically realized.

To say we may not know whether we are adding, when we are, or squaring, when we are, is actually to grant that we might perform the determinate thought function that cannot be wholly physical, and thus to grant the whole argument. Similarly, to say, "We do not know whether we ever perform a formally determinate function," is to say either (a) we are in a cognitive state, "uncertainty as to whether we are really adding, squaring, or conjoining," although we do not *experience* uncertainty, when we produce sums, squares, and simple arguments; or (b) we are always mistaken when we are certain we are adding, conjoining, etc., because at most we simulate.

Now, the first option also concedes the main argument because it postulates uncertainty when we actually do add, etc. The second postulates *mistakes* about what we are doing, and thus concedes the main argument, too: that there are such definite functions for which the only locus must be in thought. Any other answer will leave the pure function without any logical space (locus). When we are certain we are adding, we are always wrong. But that reasoning will hold for whatever we do. Thus, we are *always* wrong about what we think we are doing, when we think we are doing something definite enough to be a pure function. To suppose we can think definitely enough about functions to be wrong about what we are doing concedes the supposition of the argument again. Now the doubt has spread to include every pure function: asserting, questioning, objecting, stating, reporting, as well as adding, squaring, and conjoining. The doubt has even spread to include the very repeating of what I take, mistakenly, to be my argument and to make it indefinite whether you are actually denying or disputing my conclusion. Moreover, the cost extends to particular pure functions, specified by content: "adding three and three," "judging that Greeks are courageous," "doubting whether philosophy is scientific," "reading a paper," "thinking this writer is mad." Such an epidemic of doubt, without any effect on one's own certainty, must involve a mistake.

If we are always only simulating when we think we are doing something formally definite, then it is never determinate what we are doing at all. That requires that we are never doing such definite things at all. That *is* expensive, because there is no place for logic or mathematics or any other formal thinking at all; we cannot even "castle" in chess, but only "simulate" it, without any explanation for what "it" is or what its status is ontologically. Saint Augustine similarly objected to a "verisimilitude" account of truth in *Contra Academicos*. The relation of simulation will not be definable without the prior notion of pure functions.

If we can agree that *either* (1) we do have such definite thought processes as I described, cases of conjunction, determinate among all incompatible functions, and that they cannot wholly be physical processes (or functions among physical processes only), *or* (2) we never perform such processes but at most simulate them, then I am content. For I shall then wait for the counterattack to support (2), the one that explains the status of all those functions I cannot really perform and only think I can define (for to define one is to perform another one), and, in particular, explains the success of mathematics and pure logic, especially natural deduction systems and the proofs of completeness of propositional calculus, and offers a worked-out contrast between adding (which no one, apparently, can do) and simulating adding.

IV. FUNCTIONAL STATES

Kripke seemed to realize that functionalism would fail because "any concrete physical object can be viewed as an imperfect realization of many machine programs" (*op. cit.*, pp. 36–7, n24). But it looks to me as if he was about to draw the wrong conclusion, when he said "taking a human organism as a concrete object, what is to tell us WHICH program he should be regarded as instantiating? In particular, does he compute 'plus' or 'quus'?" He should have concluded that, if a human is only a "concrete physical object," then nothing determines, at a certain level of refinement, which program it instantiates because it instantiates none; whereas humans do add, define, and so forth, and are thus not just concrete physical objects.

If a "thought process," say, adding, were a function linking actual physical states to "subsequent" physical states, then whatever the pattern of inputs to outputs, there are incompossible functions that link the states equally well. In that case, we could not really add. Nor could we deny that we add precisely. Since we can add, we know our thought process is not the same as any function among brain states because no such function is determined (the way two points determine a line) by physical states.

The very step toward generality to escape the inconveniences of identifying an abstract process with a particular physical process, say, mechanical addition (with the inconvenience that there could be no electronic addition), creates the situation where incompossible general functions equally well "explain" the succession of physico-cognitive states, and thus discloses that no *one* function is realized to the exclusion of all the others at the physical level, and thus no pure function is realized at all. That guarantees that functions among physical states (in a process) are not the thought states because there are no determinate functions realized among physical states, when the form of thought is determinate. No real process of adding is identical with any process that equally well realizes an incompossible function. Consequently, "adding" is not a physical process or function among physical states either. Besides, the functors in such functions are not physical either. For, of course, it is numbers we add, not numerals.

V. ALL THOUGHT IS ABSTRACT

The main argument is that *some* thought is determinate, among incompossible functions, the way no physical process, series of processes, or physically determined function among processes can be. The result is that such thought is never identical with any physical process or function. (Nor can it really be such a physical process or function either, though it may, for all we have said, have a material medium, like speech.)

The full generalization that all thought is determinate that way is harder to make cogent, because it rests on one's recognizing that, whatever thinking we do, whether simple assertion or hoping or wanting or intending (over the whole family of things each of those can be, according to its particular content on a particular occasion) is such that, in order to do *that,* we have to do what is the *same* for an infinity of other cases (sorted by content) that do not happen. For someone else might have thought or said or believed or felt the same in a way definite among incompossibles. So, any thinking at all is of general "form," just as is adding, conjoining, reasoning validly, and squaring.

By its nature, thinking has "other cases" and is therefore always of a definite form (which may not be articulable by us, as are mathematical and logical forms). Asserting (in any one of its senses) cannot be "halfway" between opposed forms; it would not be asserting then. And so on, for every form of thinking. But no physical process or sequence of processes or function among processes can be definite enough to realize ("pick out") just one, uniquely, among incompossible forms. Thus, no such process can be such thinking.

The conclusion is that no physical process or sequence of processes or function among physical processes can be adding, squaring, asserting, or any other thinking at all.[17]

JAMES ROSS

University of Pennsylvania

---

[17] All thought, as content, is immaterial in two other ways. (1) It lacks the transcendent determinacy of the physical. A true judgment, "someone is knocking on my door," requires for its physical compliant reality a situation with an infinity of features not contained (or logically implied) in the true judgment. Thus, an infinity of determinate but incompossible physical situations could make the same statement true. (2) Any physical-object truth requires its truth-making reality to *overflow* the thought infinitely in the detail of what obtains. So every compliant reality is infinitely more definite then anything contingently true we can say about it. It takes a lakeful of reality for one drop of truth.

A second argument: Products of physical processes are transcendently determinate. But no product of the understanding has an infinity of content, not contained therein logically. So no physical product can ever *be* such a content of the understanding.

Some thinking is as much physical as it is immaterial. My walking, as an action, is as much a mode of thought as it is a mode of movement; yet no movement, however complex, could ever make a thought.

Leibniz says in section 17 of the *Monadology* (in *Philosophical Papers and Letters,* Leroy Loemker, ed. and trans. 2nd ed. (Dordrecht: Reidel, 1969), p. 644) that, if perception were supposed to be produced by a machine, we could make the machine on large scale and walk around in it like a mill; we would never find a perception, only the movements of wheels, gears, and pulleys. Similar reasoning is given in Leibniz's *Conversation of Philarete and Ariste* (Loemker, p. 623). I thank Margaret Wilson for pointing these passages out to me.

A third argument: The present cases concern the definiteness of the *form* of the thinking. A third, parallel argument can be constructed from the definiteness of the *content* of thought, that thought is definite among incompossible contents in a way no physical process can ever be. Similar underdetermination arguments apply.

Machines do not process numbers (though we do); they process representations (signals). Since addition is a process applicable only to numbers, machines do not add. And so on for statements, musical themes, novels, plays, and arguments.

BOOK REVIEWS

Leibniz's Metaphysics: A Historical and Comparative Study. CATHERINE WILSON. x + 350 p. Princeton: Princeton University Press, 1989. Cloth $39.50.
Leibniz and Arnauld: A Commentary on their Correspondence. ROBERT C. SLEIGH, JR. xvi + 237 p. New Haven: Yale University Press, 1990. Cloth $28.50.*

In the last ten or fifteen years, we have seen an explosion of interest in the history of philosophy in the seventeenth and eighteenth centuries among Anglo-American philosophers. We have also seen a remarkable increase in the standards of scholarship in the field. It was not so long ago that Descartes could be read as if he were an Oxford philosopher who wrote in the quaint and slightly archaic English of Haldane and Ross, and secondary literature not in English could be dismissed, as if clear-headed philosophical thought were not possible outside of English. Things are very different now. Responsible scholars must now take into account both the best available original-language texts, and the full range of secondary literature in French and German, at very least. Furthermore, work in the history of philosophy is becoming more and more historical. The scholar is expected to know not only the central philosophical texts that a Descartes or a Leibniz may have written, but the minor texts as well, texts that previously would have been dismissed as nonphilosophical and of antiquarian interest only. In addition, the well-rounded scholar is expected to know the more general intellectual and cultural background, what was going on in less visible figures like Pierre Gassendi, Sir Kenelm Digby, and Erhard Weigel, as well as parallel developments in areas like the history of science and the history of theology. Along with these changes in the field have come very healthy debates about how the history of philosophy is to be done, and what the philosophical interest of the subject is.

The new climate for the history of philosophy is very evident in these two new books, *Leibniz's Metaphysics* by Catherine Wilson and *Leibniz and Arnauld* by Robert C. Sleigh. Leibniz is a particularly difficult character to write about, in good part because he left no single text that can be regarded as definitive; the scholar of his thought must be conversant with a wide range of essays, letters, and

---

* I would like to thank Donald Rutherford, Roger Ariew, and Brandon Look for helpful comments they made on an earlier draft of this review.

0022-362X/92/8902/151–65

notes that span more than half a century, from the early 1660s to his death in 1716. This is made particularly difficult by the fact that much of the writing remains unpublished and much that is published is available only in inadequate nineteenth-century editions. Both authors show a deep acquaintance with the corpus, and both have at their fingertips the latest word in textual scholarship; indeed, Sleigh's monograph makes excellent use of some yet-unpublished manuscript work on the correspondence with Arnauld. Both also make good use of the intellectual background to Leibniz's thought, and the best of contemporary European scholarship. Furthermore, both authors are sensitive to the problem of development in Leibniz's thought. Often commentators approach Leibniz with the assumption that for any issue there is a single Leibnizian position, at least in his mature writings, and it is the job of the commentator to find it out. In reality, though, Leibniz's thought underwent a number of important changes throughout his long philosophical career; though he had his preoccupations and his commitments, he showed an admirable ability to try new things out, rethink old positions, look at his thought from different points of view, and change in the face of criticism. Both Wilson and Sleigh show a real willingness to follow Leibniz in his philosophical meanderings, and try to tease out his train of thought, rather than impose some one doctrine on him, willing or not.

Of the two books, Wilson's is the broader. The book is intended as a general investigation into Leibniz's metaphysics, "an examination of Leibniz's system understood [not] as a collection of theses and principles, but developmentally and thematically" (2). More particularly, she indicates that her main interest is the problem of the continuum, the problem as to how apparently continuous extended things in the world could be made up out of more fundamental parts. The question of historiographical methodology is also of particular concern to Wilson. Understanding Leibniz's development is central to her project, as is understanding Leibniz in his intellectual context. "What a philosopher means to say," she writes, "can only be grasped in light of the choices actually open to him, and these are defined by his relation to past and present authors" (2). Wilson's interest in development and context is evident, as is her special interest in the development of the metaphysics of the *Monadology* out of the somewhat different system(s) of the *Discourse on Metaphysics*. But the book does not really have the kind of unity that the author promises; it is, in reality, a collection of studies of various issues that relate to Leibniz's metaphysics. Each chapter is a relatively self-contained essay in an important topic in Leibniz's metaphysics which brings in both the full range of Leibniz's texts and views, and those of his contemporaries. The topics she discusses include Leibniz's interest in universal languages and encyclopedias, his thought in the 1670s before his mature period, the metaphysical systems of the *Discourse on Metaphysics,* his interest in the metaphysical foundations of physics, the development of the concept of the monad from his earlier thought, and his conceptions of space, motion, and gravity. In addition, Wilson also has chapters that focus on metaphysical issues in Leibniz's two philosophical books, the *New Essays Concerning Human Understanding* and the *Theodicy.* She ends with a very interesting and novel discussion on "the difficult problem of reading Leibniz in a post-Kantian era" (306), that is, how we can take Leibniz seriously after Kant's critique of dogmatic metaphysics.

There are many riches in Wilson's book. But what I would like to concentrate on is what she announces as her main theme, the development of Leibniz's thought.

Although development is her supposed theme, there is little to be found of it in her opening chapters, an account of Leibniz's earliest work on logic, language, and the project of an encyclopedia (chapter I), followed by an account of his metaphysics and natural philosophy in the years that precede the *Discourse on Metaphysics,* what she calls Leibniz's "first philosophy" (chapter II). Wilson's exposition here is particularly useful, given how little material about these early years is generally available in English. She is particularly good at setting the early thought about language and logic into its proper intellectual context, bringing in both familiar names like Hobbes and Locke, and less familiar views of Kircher and Wilkins. She also takes note of the important early works on physics, the *Theory of Abstract Motion* and the *Theory of Concrete Motion,* and Leibniz's so-called Paris Notes. But even though much of this early material provides an obvious contrast with Leibniz's later views (as well as obvious anticipations of his mature doctrines), Wilson has little to say about how Leibniz's thought developed in this early period, and how it may have moved from those early views to the later thought.[1]

Wilson's interest in Leibniz's development begins in earnest only in the mid 1680s with the composition of the *Discourse on Metaphysics,* an important, though neither final nor entirely coherent statement of Leibniz's first mature thought. It is in connection with the *Discourse on Metaphysics* that Wilson makes one of her most interesting claims:

---

[1] On this see, e.g., Yvon Belaval, "Premières animadversions de Leibniz sur les *Principes* de Descartes," in *Mélanges Alexandre Koyré v. II: L'aventure de l'espirit* (Paris: Hermann, 1964); Michel Fichant, "Neue Einblicke in Leibniz' Reform seiner Dynamik (1678)," *Studia Leibnitiana,* xxii (1990): 48–68; and

> To the reader who approaches it free of presuppositions, the treatise reveals not a single system, but three separate schemes or semi-systems, A, B, and C, seated in the general context of a theory of the nature and perfections of God. Metaphysics A is the theory of individual substance, based... upon Leibniz's logic of descriptions. Metaphysics B is the theory of bodies, or corporeal substances, scholastically inspired but given an extra dimension by Leibniz's theory of the *vis viva*. Metaphysics C, finally, the theory of the harmony of perceptions and actions, has strong affinities to the system of Malebranche (80).[2]

Wilson argues that these three "semisystems" are in tension with respect to one another, and do not fit together to form a single coherent system.

Although she does not put it in exactly this way, the story Wilson tells of Leibniz's development in the years that follow seems to be a story of how these different strands develop, and how the balance among them changes.

The next major stage Wilson sees in Leibniz's development occurs when he turns his attention more and more toward physics and its foundations. She writes:

> During the late 1680s, as Leibniz grew increasingly preoccupied with the metaphysical significance of empirical force-concepts, the basis of his metaphysics began to change from logic to natural philosophy (156).

Wilson's claim is that in working out his physics in the late 1680s, Leibniz came to emphasize more and more what she calls metaphys-

---

esp. André Robinet, *Architectonique disjonctive, automates systémiques et idéalité transcendentale dans l'oeuvre de G. W. Leibniz* (Paris: Vrin, 1986). These references trace carefully the emergence of a number of recognizably Leibnizian doctrines in the late 1670s. Also important here will be Christia Mercer's *Leibniz's Metaphysics* (New York: Cambridge, forthcoming), which will focus on the development of the Leibnizian conception of substance from the student work of the 1660s to the late 1670s.

[2] Although Wilson is interested in the various threads of Leibniz's thinking in the *Discourse,* she does not make use of Henri Lestienne's widely available edition of that work, which presents the complete variants found in earlier versions of that work.

ics B from the *Discourse on Metaphysics,* a metaphysics that takes as basic corporeal substance, understood as a unity of active and passive forces, which were to be identified with the form and matter of the schoolmen.

The final important change Wilson recognizes occurs in the *New System* of 1695, where, according to her, Leibniz substitutes the notion of the monad for that of the individual substance (metaphysics A) or corporeal substance (metaphysics B) as the basis of his metaphysics. Wilson argues that the monad must be carefully distinguished from the individual substance of metaphysics A of Leibniz's *Discourse*. Individual substances, as Wilson understands them, are the logical subjects of metaphysics A, those things capable of having a complete individual concept. Wilson claims that "the theory of individual substance has been dropped from the *New System*. Leibniz now comes to lean on metaphysics C . . . " (159). This is what seems to win out in the end, Wilson reports. Metaphysics A, dropped by the mid-1690s, never returns; metaphysics B, Wilson claims, was "destined for the scrap heap of 'popular notions' " (181). Metaphysics C, the theory of mind-like monads characterized by their perceptions and appetitions, the theory first unambiguously endorsed in the *New System,* is the system that predominates until the end, it seems.

Wilson's version of the development of Leibniz's metaphysics is very illuminating, and it is very healthy to see such bold interpretative hypotheses stated and defended. Wilson's Leibniz is definitely not the static character found in much of the older literature. But, at the same time, I do not think that the story she tells is quite the right one.

First of all, what Wilson identifies as metaphysics A in the *Discourse on Metaphysics* is, I think, a bit broader than the way she characterizes it, and persists longer than she allows. As she characterizes it, the focus of metaphysics A is the individual substance, characterized as Leibniz does in section 8 of the *Discourse on Metaphysics,* in terms of the predicate-in-notion principle and the complete individual concept. That is certainly important to Leibniz in the period, and it certainly is fair to say that the logical conception of this individual substance does seem to drop out. But behind the particular characterization of individuality in the *Discourse* there is a much more general concern with individuality and the closely connected notion of unity in the writings of the period. The later letters in Leibniz's correspondence with Arnauld are dominated by a concern with the importance of there being genuine individuals in the world. The argument that comes up over and over again is what might be called the aggregate argument: if there are to be complexes (aggregates) in the world (which bodies must be because of their divisibility), then there must be genuine individuals.[3] This argument persists throughout Leibniz's career, and can be seen as late as the opening sections of the *Monadology* and *Principles of Nature and Grace.*

A second reservation I have about Wilson's story concerns an aspect of Leibniz's thought in the *Discourse on Metaphysics* and beyond (not to mention before) which gets inadequate attention in Wilson's presentation of the main lines and evolution of Leibniz's thought. The denial of occasionalism, the doctrine that God is the only genuine cause in the world and the assertion of a world of genuinely active individuals, is one of Leibniz's most important themes. It is clearly present in the *Discourse on Metaphysics.* The infamous section 8 begins by noting that "it is rather difficult to distinguish the actions of God from those of his creatures,"[4] and claiming that the discussion that follows will clear this matter up; indeed, the logical theory of substance in the sections that follow can plausibly be read as an attempt not only to outline a logical notion of individuality (Wilson's metaphysics A), but to put forward a refutation of occasionalism based on undeniable logical principles. His point is that, if every individual substance has a complete individual concept, a concept that includes everything true of the individual, past, present, and future, then there must be something *in* that individual substance by virtue of which the complete individual concept pertains to it, an individual nature that answers to the complete individual concept. And if so, then we do not have to look outside of the individual (even to God) to explain any of its accidents. The refutation of occasionalism is also a central theme in Leibniz's correspondence with Arnauld, in the "Specimen Dynamicum" of

---

[3] For an account of the aggregate argument, see Daniel Garber, "Leibniz and the Foundations of Physics: The Middle Years," in K. Okruhlik and J. Brown eds., *The Natural Philosophy of Leibniz* (Dordrecht: Reidel, 1985), §I. This argument is closely connected with Leibniz's mature response to the problem of the composition of the continuum. Although announced as a theme in the beginning of the book, and although Wilson gives a serious treatment of Leibniz's early thoughts on the problem (58–60; 74–8), Leibniz's mature thought on the problem is not prominent in Wilson's exposition (see 3, 102–3, 187–8).

[4] C. I. Gerhardt, ed., *G. W. Leibniz: Die philosophischen Schriften* (Berlin: Weidmannsche Buchhandlung, 1875–90), vol. 4, p. 432, in R. Ariew and D. Garber, eds. and trans., *Leibniz: Philosophical Essays* (Indianapolis: Hackett, 1989), p. 40.

1695, in the "On Nature Itself" of 1698, and in numerous other writings.[5]

The notion of activity and the refutation of occasionalism is every bit as important as the themes that Wilson chooses to emphasize in the *Discourse on Metaphysics* and beyond. Now, Wilson does not ignore the issue of occasionalism altogether (see, e.g., 111). But because of a certain confusion in her treatment, it does not get the prominence that it deserves. Wilson closely connects the issue of God's continual recreation of the world from moment to moment, and the doctrine of occasionalism (see 148ff; 169ff.). And indeed, there is a strong temptation to do so. For many of the later Cartesians (though probably not for Descartes himself), the doctrine of God's sustenance of the world through continual recreation is the ground of their view that God is the only genuine cause in the world.[6] But Leibniz's view is different. For Leibniz, God must create the world, and must then sustain the world that he created. But what he creates and sustains is a world of active individuals. Once we separate the issues, then Leibniz's position becomes much clearer. A resolute foe of occasionalism from his earliest writings,[7] his project is figuring out how to maintain this without rejecting the doctrine of divine substance from moment to moment. But because Wilson does not carefully separate the issue of divine sustenance from occasionalism, she takes Leibniz's commitment to divine sustenance as evidence of his commitment to occasionalism. Her view seems to be that Leibniz does not really reject occasionalism in the *Discourse on Metaphysics*, and argues that even in later writings (as late as the *Monadology* in her view) Leibniz is not sure on that issue, a position that cannot be right (see 169–70).

If we read metaphysics A in an appropriately broad way, and add activity to Wilson's three strands of Leibniz's metaphysics, then we have, I think, a reasonably good list of Leibniz's main metaphysical preoccupations. But I think that it is not entirely accurate to see these themes simply as competing semisystems, and to see Leibniz's development as simply a matter of choosing among them. Wilson is not altogether wrong. Leibniz's metaphysics is grounded in the idea that there are certain metaphysically basic substances that are active and genuine unities, and at different times he does hold different conceptions of what those active individuals are. On one model, the metaphysically basic active individuals are corporeal substances, thought of on analogy with animals (Wilson's metaphysics B); on another, they are conceived of on analogy with Cartesian souls (Wilson's metaphysics C). (There may be others too.) Hence the apparent conflict between metaphysics B and metaphysics C which Wilson notes. Although both models of what is metaphysically basic remain live possibilities for much of his life, there is a general progression in Leibniz's thought from the first (most prominent in the letters to Arnauld in the late 1680s) to the second (most prominent in the *Monadology*). But even when Leibniz chooses to ground his metaphysics in mind-like monads, corporeal substance does not necessarily drop out; to this extent, metaphysics B and C are not really competitors. Although they may not be metaphysically basic (as they seem to be in the letters to Arnauld), corporeal substances are often present in Leibniz's later philosophy; this, indeed, is a central problem in the later part of Leibniz's correspondence with Des Bosses, how corporeal substance could be possible in a world of monads. To make sense of Leibniz's thought, to appreciate both what remains constant and what changes, I think that we must distinguish between at least two levels in Leibniz's metaphysics. There is, first of all, what we may call the abstract metaphysics, basic, structural principles. Here I would include the commitment to a world of metaphysically basic genuine individuals, individuals that are genuinely active and ground the rest of the contents of Leibniz's world. It is here that we find the emphasis on individuality characteristic of Wilson's metaphysics A, which is not in competition with her metaphysics B or C, but grounds them. This abstract metaphysics seems to change little in its broad outlines in his mature thought. But that settled, there is a great deal that is not. What remains is what might be called the concrete metaphysics, the account of what exactly it is that is metaphysically basic (are they mind-like or animal-like or something else altogether?), and how exactly the metaphysically basic grounds the rest of the furniture in Leibniz's world.

---

[5] In the correspondence with Arnauld, see, e.g., Gerhardt, vol. 2, pp. 92–4, in Ariew and Garber, pp. 82–4; in the "Specimen Dynamicum," see Gerhardt, ed., *G. W. Leibniz: Mathematische Schriften* (Berlin: Asher & Halle: Schmitt, 1849–63), vol. 6, p. 242, in Ariew and Garber, p. 125; in "On Nature Itself," see §§ 8–10, Gerhardt, vol. 4, pp. 508–10, in Ariew and Garber, pp. 159–61.

[6] See Garber, "How God Causes Motion: Descartes, Divine Sustenance, and Occasionalism," this JOURNAL, LXXXIV, 10 (October 1987): 567–80.

[7] For Leibniz's first clear rejection of occasionalism, see Garber, "Motion and Metaphysics in the Young Leibniz," in M. Hooker, ed., *Leibniz: Critical and Interpretative Essays* (Minneapolis: Minnesota, UP, 1982), esp. pp. 166–7, 171–2. This is also discussed in Mercer's forthcoming monograph.

Wilson's book is a real contribution to the English-language literature on Leibniz. But, nevertheless, it contains a number of straightforward mistakes. Mersenne's *Harmonie universelle* is 1636–37, not 1619 (23); Leibniz's "Brief Demonstration" of 1686 is twice given as 1684 (129, 138); the reference to Leibniz's view that God is a concurrent mind that gives nonhuman substances their reality is surely not to anything in 1678, as Wilson claims (183), but to writings that date from 1668.[8] In one other place, Wilson uses a very early document, the *Propositions quaedam physicae* of 1672 to support a claim about Leibniz's mature views of space, attributing to Leibniz a doctrine unlikely for him to have held in his later thought (134). In another, she identifies the quantity conserved in Descartes's universe as "weight x speed" (142), rather than size times speed.[9]

Although Sleigh is more interested in the philosophical analysis of Leibniz's thought than Wilson is, his book shares the ambition of explicating Leibniz's views in their proper historical context. But rather than take the long view, Sleigh concentrates on a very few texts; while he has an evident command of the breadth of Leibniz's writings, the focus is on the *Discourse on Metaphysics* and the important correspondence with Arnauld. Sleigh attempts to present a picture of what the world looked like to Leibniz at that crucial moment in his development, in 1686, what Leibniz himself identified as the beginning of his mature thought. The result is an exemplary piece of scholarship, historically sound and philosophically illuminating, a model study of how to treat the elusive thought of a character like Leibniz.

After some methodological preliminaries, Sleigh offers some useful background on two figures essential to understanding the historical place of the writings he is studying. There is a very informative chapter on "the Great Arnauld," Antoine Arnauld, the recipient of Leibniz's letters, and the author of some very penetrating comments on Leibniz's ideas (chapter 3). Sleigh there sketches out Arnauld's full career and outlines his views on key philosophical and theological issues relevant to his exchange with Leibniz. In addition, Sleigh has a short chapter (chapter 2) on the intermediary in the correspondence, Ernst, Landgrave of Hessen-Rheinfels, a Catholic nobleman sympathetic to Leibniz. (It was also no small matter that Ernst happened to know how to get in touch with the Great Arnauld.)

---

[8] The reference is, presumably, to the sort of account Leibniz gives in the "De Transubstantione" of 1668, in the Akademie edition, *G. W. Leibniz: Sämtliche Schriften und Briefe* (Berlin: Akademie, 1923–), series 6, vol. 1, pp. 508–9, in Leroy Loemker, ed. and trans., *Leibniz: Philosophical Papers and Letters* (Dordrecht: Reidel, 1969), p. 116.

[9] Weight, of course, presupposes gravity (heaviness), which for Descartes was a derivative physical magnitude, one that bodies have only because of the fluid vortex that happens to surround the earth.

But the heart of the book is a series of four chapters on Leibniz's philosophical thought in the period. In chapter 4, Sleigh discusses the issues of freedom and contingency in the texts, and the related issue as to whether all properties an individual has are essential to it (the so-called doctrine of superessentialism). The issues here are familiar ones, some of the most widely discussed of Leibniz's views, at least in the English-language literature. Sleigh's discussion of them is very skillful, showing his mastery of the texts, the contemporary literature, and the philosophical issues. And despite the familiarity of the issues in recent discussions, Sleigh manages to make some surprising and original claims, arguing against superessentialism and for a weaker doctrine he calls superintrinsicalness. In chapter 5, Sleigh takes up the question of substance in the late 1680s, showing the variety of views about what the metaphysically basic substances are among which Leibniz was attempting to decide, whether such basic substances are to be understood on the model of a soul or mind, or on the model of some sort of organism, a question to which we shall return. Chapter 6 is entitled "Substantial Forms." While the notion of a substantial form is certainly relevant to the discussion there, the main focus is on the question of the unity of substance for Leibniz, how it is that the various candidates for substancehood that Sleigh discusses in chapter 5 constitute genuine individuals and unities. Chapter 7 ("Action") is a lovely account of Leibniz's views on the activity of substance. Sleigh is very good here at setting out the different views of Leibniz, Arnauld, and Malebranche (whose views are behind so much of the text of the *Discourse*) on the question of occasionalism and the action of God and his creatures. (Sandwiched in here is also an elegant, though brief, discussion of expression, accommodation, and solipsism in Leibniz (170–82).) The book ends with a brief epilogue (chapter 8, "Concluding Remarks"). Among the appendices is a translation of a draft of the short letter of 14 July 1686 (one of two letters written on that date), appearing in print for the first time in any language. It is worth remarking that Sleigh uses the very latest in scholarly tools. For the *Discourse on Metaphysics,* he uses the edition of Lestienne, containing all of the variants from all extant versions of the text.[10]

---

[10] Unfortunately, the Academy edition published its own preliminary version of that text in its *Vorausedition* only after Sleigh's manuscript was complete; it appears in the *Vorausedition,* vol. 8 (1989), pp. 1687–1739.

For the correspondence with Arnauld he makes use of the edition that Geneviève Rodis-Lewis published of the letters that Arnauld actually received,[11] as well as the new edition of the Hanover manuscripts in preparation by Reinhard Finster, based on a completely new transcription of the manuscripts and containing a significant amount of philosophically important material not currently available in published form.

There is so much to discuss in Sleigh's clear and elegant book. But it would be interesting to discuss one theme that Sleigh's book has in common with Wilson's, Leibniz's views on substance, body, and the ultimate make-up of the world in the period in question.

Sleigh begins by distinguishing four different accounts of what the metaphysically basic substances in the world are like, what he calls the spiritual theory, the corporeal-substance theory, the modified-corporeal-substance theory, and the monadological theory (96). On the spiritual theory, all there are in the world are spirits, that is, conscious minds (ours, God's, and those of angels); bodies on this view are treated phenomenalistically, as bundles of (conscious) perceptions (98). On the corporeal-substance theory, the metaphysically basic elements of the world are corporeal substances, unities of form and matter on analogy with organisms, where form and matter are themselves understood not to be substances capable of existing apart from the composite (98–9). The modified-corporeal-substance view is just like the straight view, except for the fact that the form (soul) is itself a substance, capable of independent existence, even if it never actually exists without a body (99–100). On either view, the basic model for the individual is an animal, whose soul is a form, and body, matter. Finally, on the monadological theory, the metaphysically basic things of the world are monads, unextended, simple, soul-like entities, only some of which are conscious (100–1).

Like Wilson, Sleigh sees the *Discourse on Metaphysics* and the subsequent correspondence with Arnauld as a combination of different threads, not surprisingly, the four threads enumerated above. At the surface, most visibly in the correspondence with Arnauld, is the corporeal-substance theory, particularly in its modified form, in which the form (soul) is regarded as something of a substance (109). But often lurking beneath the surface of the *Discourse* and correspondence with Arnauld, in somewhat elliptical passages and in passages that Leibniz deleted from his final drafts, as sent to Arnauld,

---

[11] Geneviève (Rodis-) Lewis, ed., *Lettres de Leibniz à Arnauld d'après un manuscrit inédit* (Paris: Presses Universitaires de France, 1952).

Sleigh sees evidence of both the spiritual theory and the monadological view. One feature that Sleigh calls our attention to is the numerous times, both in deleted passages and in passages which Leibniz allowed to stand, in which he treats the reality of body hypothetically: *if* bodies are real, *if* bodies are not mere phenomena like the rainbow, ***then*** the world must contain substances that have such-and-such properties. Sleigh takes Leibniz's apparent agnosticism about body to be a sign that he is attempting to turn Arnauld's attention to the spiritual or monadological theory (103, 104; cf. 105). He does acknowledge that the corporeal-substance view, in one or another of its variants, does seem to win out, particularly in the correspondence with Arnauld (109); in this respect he seems to differ from Wilson, for whom the corporeal-substance view does not predominate until later in the decade, as Leibniz turns his attention to physics. But despite the visibility of the corporeal-substance view, he does not think that it is Leibniz's real view in these texts. Sleigh is convinced by certain arguments that Leibniz offered against the reality of motion and extension in these and related texts that the corporeal-substance view cannot represent Leibniz's real view; for if there are really corporeal substances, be they basic or not, they would have to be extended, and if they are extended, then, Sleigh argues, that would cause serious problems on Leibniz's view, problems Leibniz explicitly recognized (112ff). Sleigh's view seems to be that Leibniz presses the notion of corporeal substance in his correspondence with Arnauld not because he believes it, but only to show that his philosophy is suitable for grounding Catholic theology, making it more suitable than Descartes's view for unifying the Catholic and Protestant churches, a project dear to Leibniz's heart. Sleigh notes that Leibniz calls Arnauld's attention to the official Catholic dogma that the soul is the form of the human body, something that Leibniz can say on his own corporeal substance view, but something that may be more problematic on Descartes's brand of dualism (108; cf. 115). In his heart, Sleigh implies, Leibniz actually preferred the monadological view.

Sleigh's account of Leibniz's thought is certainly very illuminating, and based on very careful reading of the sources. But I am not altogether convinced. First of all, I see problems with Sleigh's attribution of the spiritual view to Leibniz in these texts. Leibniz certainly toyed with the view that the basic constituents of his world are like souls; but I see very little reason to believe that, in the period in question, he held that the only such souls were spirits in the sense that Sleigh uses the term, conscious souls, capable of reason, as opposed to what he was later to call monads, which come in all varieties, conscious and unconscious. In support of the spiritual view, Sleigh points to two sections of the *Discourse*, 14 and 34, and to a deleted passage from section 8 (98, 101). But section 14 begins with passage in which Leibniz quite explicitly argues that there is a substance that corresponds to *every* point of view God has on the world, something unlikely if the only substances are spirits. And section 34 begins with an explicit distinction between intelligent souls and souls that "do not know what they are nor what they do."[12] The best evidence Sleigh presents is a deletion Leibniz made from an early version of section 8 of the *Discourse*. In that version, Leibniz used the ring of Gyges as an example of an individual substance. In that version, Leibniz originally wrote: "I speak here as if it were assured that this ring had a consciousness" (quoted 101). This, indeed, shows that Leibniz *may* once have regarded being conscious as necessary for substancehood. But, on the other hand, he did cross this passage out, and replace the phrase 'had a consciousness' with 'is a substance', as Sleigh notes. He might (quickly) have realized that, while consciousness might be a sufficient condition for constituting the ring a genuine substance, it is not necessary; a lesser soul or form would do just as well. Or it might have been an absent-minded slip of the pen. In either case, it seems slim evidence to ground the claim that Leibniz took seriously Sleigh's spiritual view at this stage of his thought.

Secondly, I am not convinced that the Leibnizian critiques of the reality of extension Sleigh cites are that damaging to a corporeal-substance point of view. To be sure, if Leibniz believes in the reality of corporeal substance (not to mention the claim that they are metaphysically basic, as he suggests in the correspondence with Arnauld), then he is admitting something into his world that is extended in a way that minds, souls, or monads are not. But that is not to say that corporeal substances are extended in the deepest sense. Leibniz can certainly hold (and probably did hold, on my view) that what is basic in body is force, and that extension is ultimately to be understood in terms of the forces bodies have. In particular, Leibniz seems to hold that extension results from the primitive passive force that excludes the penetration of one body by another. On this view, an extended body can be regarded as the diffusion of the force of impenetrability that defines a region that is the space occupied by a given body. Extension is, at least to that extent, less basic, metaphysically speaking, and thus less real than force. Although on Sleigh's reading the arguments criticizing the notion of extension lead to a monadological metaphysics, in which extended things are entirely phenomenal, on my reading these same arguments are really part of his support for this somewhat different position, a world of corporeal substances grounded in active and passive force.

---

[12] Sleigh, in fact, cites (98) this distinction in his account of the spiritual theory. He also notes (102) that the autograph begins with a passage, subsequently deleted, in which Leibniz says that he does not undertake to show whether or not there are "substances, souls, or substantial forms that are not intelligent." While this is perfectly consistent with the spiritual view, it is equally consistent with its contrary.

Consider, for example, the argument which Sleigh especially emphasizes, in which Leibniz argues that any body (corporeal substance) is made up of an actual infinity of smaller parts, from which he concludes not merely that they have shapes of infinite complexity, but also that, in the strictest sense, they have no shape at all (112–5). Both Sleigh and I could agree that Leibniz's intention here is to undermine the Cartesian claim that the essence of body is extension, that extension is something that characterizes body in its innermost nature. Anything really extended, essentially or otherwise, must have some shape or another; so, by showing that despite appearances, bodies do not *really* have shapes, Leibniz shows that they are not really *extended* either. This *may* lead one to a world of nonextended and mind-like monads. But it might also be used as an argument against the existence of Cartesian substances whose essence is extension, and toward the sort of view I sketched above, on which what is basic is force. While arguments against the reality of extension are intended to lead one beyond the kind of superficial account of body one finds in the Cartesian view, they do not really tell us where we hit bottom, metaphysically speaking, whether it is the corporeal substances with their active and passive forces, or the monads with their perceptions and appetitions that are supposed to ground the world of phenomena.

Sleigh acknowledges this in a way, and admits that an advocate of corporeal substance in Leibniz, like myself, can read this argument in this manner. In the end, his reading of the argument rests on the claim that "lurking in the background" is either a monadological analysis of body, or, perhaps weaker, the idea that there must be some analysis of that sort to be found and developed (115). But this is to admit that the issue is not the argument itself, but what it is that is "lurking in the background." And whatever may be lurking in the background, it seems clear that, when writing to Arnauld, what is present in the foreground is a metaphysical picture grounded in the notion of corporeal substance. I, for one, see no reason to prefer the view hidden in the shadows to that which Leibniz presents explicitly. In the end, I much prefer to think that Leibniz stressed the corporeal-substance view when writing to Arnauld not for any religious or political reasons, but because it is, quite simply, what he believed, not always before or always after, but at that time. One can certainly find monadological currents in Leibniz's thought in these years. But until someone can convince me that these undercurrents and suggestions are more representative of Leibniz's true thoughts than what Sleigh admits Leibniz wrote to Arnauld, I shall continue to read Leibniz as writing what he meant, not only here, but in the many other texts of the period in which the corporeal-substance view is emphasized.[13]

Although Sleigh and I may disagree on this point, this in no way decreases my admiration for his accomplishment; despite our disagreements, his book remains, in my estimation, a brilliant piece of historical scholarship and philosophical analysis. His book and Wilson's are marks of how far scholarship in the history of philosophy has come over the last few years, models of the best in contemporary scholarship. One can only hope that they will have the influence on the field that they genuinely deserve, and inspire others to approach the history of philosophy with the kind of care and intelligence that these books demonstrate.

DANIEL GARBER

University of Chicago

---

[13] See Garber, "Leibniz and the Foundations of Physics: The Middle Years," pp. 66–72, for an account of such texts.

NOTE TO CONTRIBUTORS

As a monthly periodical, THE JOURNAL OF PHILOSOPHY encourages philosophical exchange and progress through prompt publication of articles of current interest. The editors now feel that they can ensure publication within six to twelve months of acceptance. They continue to be particularly interested in papers that explore the borderline between philosophy and the special disciplines, but are unlikely to accept articles primarily expository or historical in character. For the Comments and Criticism section they accept rejoinders but prefer those designed to advance the discussion rather than just to set the record straight. If and only if articles are held five months or more and then rejected, letters of rejection include a brief comment concerning the editors' reasons for rejection.

Please consult the memo to authors and the statement of copyright policy, which are available from the JOURNAL office.

ERRATUM NOTICE

The following "Predation Matrix" was incomplete on page 719 of Peter Godfrey-Smith's "Signal, Decision, Action," which appeared in our December 1991 issue, pp. 709–22.

Predation Matrix

	S	N
yes	−2	−2
no	−42	0

S = predator in the vicinity; N = noise
yes = flight or concealment response; no = no behavior

Cambridge University Press

New in paperback…

Spectacles and Predicaments

Essays in Social Theory

Ernest Gellner

Preface by I. C. Jarvie

"…Gellner's erudite analysis of our knowledge and interpretation of reality is constantly pertinent and provocative." —*Psychological Medicine*

In this collection of nineteen essays Ernest Gellner considers what professional philosophers have to say about everyday life and society. As ever, he begins as a critic, expounds alternative views of his own, and applies them to current issues.

22486-1 Hardcover $54 50
42434-8 Paper $19.95

Beliefs in Action

Economic Philosophy and Social Change

Eduardo Giannetti da Fonseca

This book is concerned with the role of economic philosophy ("ideas") in the processes of belief-formation and social change. Its aim is to further our understanding of the behavior of the individual economic agent by bringing to light and examining the function of non-rational dispositions and motivations ("passions") in the determination of the agent's beliefs and goals.

39306-X Hardcover $39 95

Interpersonal Comparisons of Well-Being

Jon Elster and John E. Roemer, Editors

In this volume a diverse group of economists, philosophers, political scientists, and psychologists address the problems, principles, and practices involved in comparing the well-being of different individuals. This collection constitutes the most advanced and comprehensive treatment of one of the cardinal issues in social theory.

Contributors: *T. M. Scanlon, James Griffin, John Broome, Jon Elster, Daniel Kahneman, Carol Varey, Charles Blackorby, David Donaldson, Peter J. Hammond, John Weymark, Ignacio Ortuno-Ortin, John E. Roemer, Aanund Hylland, Claude d'Aspremont, L. A. Gerard-Varet.*

Studies in Rationality and Social Change

39274-8 Hardcover $49 50

Political Theory and Postmodernism

Stephen K. White

This book outlines a path through the postmodern problematic by distinguishing two distinct ways of thinking about the meaning of responsibility, one prevalent in modern and the other in postmodern perspectives. Using this as a guide, White explores the work of Heidegger, Foucault, Derrida, Lyotard, and Habermas, as well as "difference" feminists, with the goal of showing how postmodernism can inform contemporary ethical-political reflection.

Modern European Philosophy

40122-4 Hardcover $39.50
40948-9 Paper $12 95

Liberal Purposes

Goods, Virtues, and Diversity in the Liberal State

William A. Galston

This book is a major contribution to the current theory of liberalism by an eminent political theorist. It challenges the views of such theorists as Rawls, Dworkin, and Ackerman who believe that the essence of liberalism is that it should remain neutral concerning different ways of life and individual conceptions of what is good or valuable.

Cambridge Studies in Philosophy and Public Policy

41036-3 Hardcover $49 50
42250-7 Paper $16 95

Violence, Terrorism, and Justice

R. G. Frey and Christopher W. Morris, Editors

This volume addresses the urgent problem of terrorism. The essays define terrorism, discuss whether the assessment of terrorist violence should be based on its consequences (beneficial or otherwise), and explore what means may be used to combat those who use violence without justification.

Contributors: *Annette C. Baier, Claudia Card, Jonathan Glover, Virginia Held, Thomas E. Hill, Jr., Gregory S. Kavka, Loren E. Lomasky, Jan Narveson, Onora O'Neill, Alan Ryan, Thomas C. Schelling, R. G. Frey, Christopher W. Morris.*

Cambridge Studies in Philosophy and Public Policy

40126-9 Hardcover $54 50
40960-0 Paper $17 95

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.
MasterCard/VISA accepted. Prices subject to change.

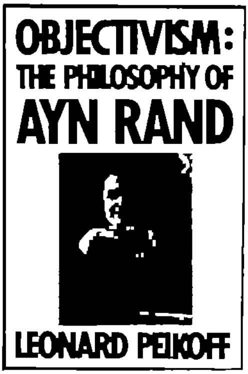
From Penguin Books USA

Objectivism: The Philosophy of Ayn Rand

by Leonard Peikoff

The definitive statement of Ayn Rand's philosophic system, written by her foremost interpreter

To the general public, Ayn Rand is best known for her defense of egoism and individualism, but her most radical and original contributions lie in basic philosophy—particularly in her theory of concept-formation and in her thoroughly non-Kantian theory of the relationship of consciousness to existence.



This 500-page work is based on a 1976 lecture course that Peikoff prepared with the assistance of Rand herself. Rand described this course as "the only *authorized* presentation of the entire theoretical structure of Objectivism, i.e., the only one that I know of my own knowledge to be fully accurate."

Leonard Peikoff was Rand's close associate from the early 1950's until her death in 1982 and is recognized as the leading exponent of Objectivism. Peikoff has taught philosophy at New York University, Long Island University, Hunter College, and the University of Denver. He is the author of *The Ominous Parallels* (a study of the philosophic roots of Nazism), and of numerous articles on Objectivist philosophy, including "The Analytic-Synthetic Dichotomy," which Rand included in her work *Introduction to Objectivist Epistemology*.

Objectivism: The Philosophy of Ayn Rand covers such topics as:

- **Existence, Consciousness, and Identity as the basic axioms**
- **Causality as a corollary of Identity**
- **Existence as possessing primacy over consciousness**
- **The senses as necessarily valid**
- **Human actions as both caused and free**
- **Concept-formation as a mathematical process**
- **Concepts as objective**
- **Knowledge as contextual and hierarchical**
- **The arbitrary as neither true nor false**
- **Reason as man's basic means of survival**
- **Man's life as the standard of moral value**
- **Rationality as the primary virtue**
- **Justice as rationality in the evaluation of men**
- **The initiation of physical force as evil**
- **Individual rights as absolutes**
- **Capitalism as the only moral social system**
- **Art as a concretization of metaphysics**

FOR A COMPLIMENTARY COPY, WRITE TO:

The Ayn Rand Institute, Dept. JNP
P.O. Box 6004, Inglewood, CA 90312

COMMON KNOWLEDGE

A New Journal from Oxford University Press

From the Inaugural Volume

Calls For Papers

Stephen Toulmin

Clifford Geertz

Natalie Zemon Davis

Elizabeth Fox-Genovese

Frank Kermode

William M. Chace

Stanley N. Katz

G. Thomas Tanselle

Cornelius Castoriadis

Stanley Cavell

Steven Marcus

Mikhail Epstein

Greil Marcus

Paul Feyerabend

Eugene Genovese

Jeffrey M. Perl

Columns

Richard Rorty

Adam Michnik

W. V. Quine

Miroslav Holub

Vicki Hearne

Paul Feyerabend

Articles

Left-Wing Wittgenstein,
Right-Wing Marx
Bernard Williams

Toward a Philosophy of Style
Manfred Frank

The Achievement of Ambivalence
Hanna Segal

Feminism and the Rhetoric of Individual Rights
Elizabeth Fox-Genovese

Sixty-Eight Eighty-Sixed
Greil Marcus

Books, Canons, and the Nature of Dispute
G. Thomas Tanselle

Poetry/Fiction

The Faithful Lover and four other poems
Denise Levertov

What Crime?
(chapters from a novel)
Julia Kristeva

Fifty-Six Guinea Pigs over the Garden
(a novella)
Péter Esterházy

Reviews

Conversations with Claude Lévi-Strauss,
edited by Didier Eribon
Clifford Geertz
Frederick Turner
Simon Goodhill

Delayed Book Review: The Nature of Peace
(1916) by Thorstein Veblen
Sidney Plotkin

Briefly Noted
Caroline Walker Bynum
Paul Feyerabend
Joseph Frank

The Collected Poems of Henri Coulette
Philip Levine

Unpublished Notebooks
(selected and with an introduction by Donald
Justice and Robert Mezey)
Henri Coulette

To subscribe for 1992 (3 issues)
at the special charter rate,
send check or money order
for $19 to the
Journals Department,
Oxford University Press,
2001 Evans Road,
Cary, NC 27513,
or call toll free 800-852-7323

Oxford

In My Father's House

Africa in the Philosophy of Culture

KWAME ANTHONY APPIAH, *Harvard University*

"This is an absorbing and path-breaking book by a gifted philosopher. Appiah rescues the philosophy of culture from Herder by insisting that we drop notions like 'authentic negritude' and that 'African culture' is the name of an important project rather than of an available datum. The book's range of reference and the vigor of its argumentation are equally impressive."—Richard Rorty, *University of Virginia*

1992 232 pp. $29.95

The Nature of All Being

A Study of Wittgenstein's Modal Atomism

RAYMOND BRADLEY, *Simon Fraser University*

"Bradley knows the historical literature extremely well....The writing is clear throughout....This book will make a significant contribution to both Wittgenstein scholarship and contemporary philosophy of logic."—Robert J. Fogelin, *Dartmouth College*

1992 304 pp. $39.95

Essays on Aristotle's *De Anima*

Edited by MARTHA C. NUSSBAUM, *Brown University*, and AMELIE OKSENBERG RORTY, *Radcliffe and Mount Holyoke Colleges*

Bringing together a group of outstanding new essays on Aristotle's *De Anima*, this book covers topics such as the relation between soul and body, sense-perception, imagination, memory, desire, and thought, which present the philosophical substance of Aristotle's views to the modern reader.

May 1992 384 pp. $69.00

✦New in paperback!

Constructibility and Mathematical Existence

CHARLES S. CHIHARA, *University of California, Berkeley*

Chihara here develops a mathematical system in which there are no existence assertions but only assertions of the constructibility of certain sorts of things and utilizes this system in the analysis of the nature of mathematics.

1990 (paper 1991) 304 pp.
paper $24.95 cloth $49.95

The Imaginary Museum of Musical Works

LYDIA GOEHR, *Wesleyan University*

Combining philosophical and historical methods of enquiry, Goehr describes how the concept of a musical work emerged as late as 1800, and how it subsequently defined the norms, expectations, and behavior characteristic of classical musical practice.

May 1992 320 pp. $65.00

Drawing the Line

Life, Death, and Ethical Choices in an American Hospital

SAMUEL GOROVITZ, *Syracuse University*

"A small gem of medical philosophy....Captivating."—*Kirkus Reviews*. "This book could have a definite impact on the thinking and planning of all those associated with health care."—*Booklist*.

1991 224 pp. $19.95

Springs of Action

Understanding Intentional Behavior

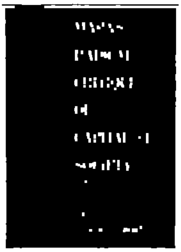
ALFRED R. MELE, *Davidson College*

Tackling some central problems in the philosophy of action, Mele constructs an explanatory model for intentional behavior, locating the place and significance of such mental phenomena as beliefs, desires, reason, and intentions in the etiology of intentional action.

1991 288 pp. $37.95

✦*New in paperback!*

Marx's Radical Critique of Capitalist Society



A Reconstruction and Critical Evaluation

N. SCOTT ARNOLD, *University of Alabama, Birmingham*

In this book, Marx's charges of exploitation and alienation against capitalism are explicated and evaluated as part of a radical critique of capital society.

1990 (paper 1991) 352 pp.
paper $14.95 cloth $38.00

The New Art of Autobiography

An Essay on the Life of Giambattista Vico Written by Himself

DONALD PHILLIP VERENE, *Emory University*

In this, the first full-length study of Vico's highly original autobiography, Verene discusses its place in the history of autobiography generally, and shows it to be the first work of modern intellectual autobiography which uses a genetic method.

1991 280 pp. $59.00

The Continuous and the Discrete

Ancient Physical Theories from a Contemporary Perspective

MICHAEL J. WHITE, *Arizona State University*

Discussing three ancient models of spatial magnitude, time, and local motion, this book places them within the context of later philosophical, scientific, and mathematical developments.

April 1992 340 pp. $69.00

The Conception of Value

PAUL GRICE

Introduction by JUDITH BAKER

The works of Paul Grice collected in this volume present his metaphysical defense of value, and represent a modern attempt to provide a metaphysical foundation for value. The most striking part of Grice's position is his contention that the legitimacy of such evaluations rests ultimately on an argument for absolute value.

1991 176 pp. $45.00

Plato on the Self-Predication of Forms

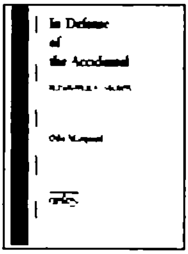
Early and Middle Dialogues

JOHN MALCOLM, *University of California, Davis*

In this book, Malcolm presents a new and radical interpretation of Plato's earlier dialogues. He argues that the few cases of self-predication contained therein are acceptable simply as statements concerning universals, and that therefore Plato is not vulnerable in these cases to the Third Man Argument.

1991 240 pp. $55.00

In Defense of the Accidental



Philosophical Studies

ODO MARQUARD, *University of Giessen, Germany*

Translated by ROBERT M. WALLACE

Ranging from the equivalence of modernity and the *theodicee* to the notion of world history, the seven essays collected in this book address the history of modern ideas and contemporary cultural issues.

1991 144 pp.
paper $12.95 cloth $32.50

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MG
To order by phone using major credit cards please call (212)679-7300, ext.7106

Oxford University Press
200 Madison Avenue ✦ New York, NY 10016

Classics in Philosophy

GREEK PHILOSOPHY
Thales to Aristotle
Third Edition, Revised and Expanded
Reginald E. Allen
1991 0-02-900495-0 $14.95 paper

POLITICAL IDEAS OF ST. THOMAS AQUINAS
St. Thomas Aquinas
0-02-840380-0 $10.95 paper

EIGHTEENTH-CENTURY PHILOSOPHY
Lewis W. Beck, Editor
0-02-902100-6 $12.95 paper

A MODERN INTRODUCTION TO PHILOSOPHY, 3rd Edition
Paul Edwards & Arthur Pap, Editors
0-02-909200-0 $27.95

A SHORT HISTORY OF CHINESE PHILOSOPHY
Fung Yu-Lan
0-02-910980-9 $14.95 paper

NINETEENTH-CENTURY PHILOSOPHY
Patrick Gardiner, Editor
0-02-911220-6 $12.95 paper

INTRODUCTORY READINGS IN AESTHETICS
John Hospers
0-02-915260-7 $14.95 paper

MORAL AND POLITICAL PHILOSOPHY
David Hume
0-02-846170-3 $12.95 paper

DIALOGUES CONCERNING NATURAL RELIGION
David Hume
0-02-846180-0 $10.95 paper

PHILOSOPHY OF THE SIXTEENTH AND SEVENTEENTH CENTURIES
Richard H. Popkin
0-02-925490-6 $12.95 paper

ESSAYS IN PRAGMATISM
William James
0-02-847140-7 $9.95 paper

CRITIQUE OF JUDGEMENT
Immanuel Kant
0-02-847500-3 $12.95 paper

PHILOSOPHY AND TECHNOLOGY
Carl Mitcham & Robert Mackey, Editors
0-02-921430-0 $14.95 paper

THE SPIRIT OF THE LAWS
Baron de Montesquieu
0-02-849270-6 $13.95 paper

NEWTON'S PHILOSOPHY OF NATURE
Isaac Newton
0-02-849700-7 $12.95 paper

A CRITICAL HISTORY OF WESTERN PHILOSOPHY
D.J. O'Connor
0-02-923840-4 $14.95 paper

GREEK AND ROMAN PHILOSOPHY AFTER ARISTOTLE
Jason L. Saunders
0-02-927730-2 $14.95 paper

ETHICS
Benedict de Spinoza
0-02-852650-3 $10.95 paper

TWENTIETH CENTURY PHILOSOPHY
The Analytic Tradition
Morris Weitz
0-02-934990-7 $12.95 paper

MEDIEVAL PHILOSOPHY
John F. Wippel & Allan B. Wolter
0-02-935630-4 $12.95 paper

Attention College Professors: Examination copies will be sent on a 30-day approval basis only. Simply place your request on university stationery, indicating book requested, course title, projected enrollment, and date your next text decision is due. Mail to

Fp **THE FREE PRESS**
A Division of Macmillan/A Maxwell Macmillan Company
866 Third Avenue, New York, NY 10022

Apr

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 4
APRIL 1992

page
167 *Why Knowledge Is Merely True Belief* Crispin Sartwell
181 *A Sentential Theory of Propositional Attitudes* Michel Seymour

BOOK REVIEWS
202 Susan Wolf: *Freedom within Reason* Bernard Berofsky

209 NOTES AND NEWS
210 NEW BOOKS
216 NEW BOOKS: ANTHOLOGIES
219 NEW BOOKS: PAPERBACKS
220 NEW BOOKS: TRANSLATIONS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027 FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 4, APRIL 1992

WHY KNOWLEDGE IS MERELY TRUE BELIEF*

I claim that propositional knowledge is merely true belief, that an agent knows that *p* if and only if *p* is true and the agent believes that *p*. Elsewhere, I have attempted to answer the obvious objections to that position.[1] Here, I want to offer a positive argument for it.

I. TERMINOLOGICAL GROUNDWORK

One objection that has been raised to the claim that knowledge is merely true belief is that it seems to leave us without any preanalytic sense of what knowledge is. That is, it seems to be a question whether the proffered analysis could possibly be an analysis of *knowledge* at all.[2] This objection seems to arise because there is a very widely held intuition that belief and truth could not possibly be sufficient for knowledge. Then the question of whether the analysis embodies a theory of knowledge seems to come down to a clash of intuitions about what notion it is which is being analyzed. For this reason, I want to offer a way of picking out the notion to be defined. Here is my proposal: *knowledge is our epistemic goal in the generation of particular propositional beliefs*. It will be convenient to have a term to refer to procedures that have as their goal the genera-

---

[*] I would like to thank Jeffrey Tlumak, who commented elaborately on an earlier version of this paper.

[1] "Knowledge is Merely True Belief," *American Philosophical Quarterly*, XXVIII, 2 (April 1991): 157–65. That essay merely lays the groundwork for a defense of the position; roughly, I try to show there that the claim that knowledge is merely true belief cannot be refuted by the flick of a counterexample. I refer those who think it can be so refuted to the earlier paper. I intended in that essay merely to reply to the claim that our ordinary use of the term 'knowledge' committed us to a distinction between knowledge and mere true belief. This appeared to me to be the only explicit argument in the literature to the effect that at least some third condition was required, and thus I hoped, by attacking that argument, to suspend the question of burden of proof.

[2] This objection was brought to my attention by Tim McGrew.

0022-362X/92/8904/167–80

tion of belief with regard to particular propositions. Let us refer to such procedures collectively as *inquiry*. Knowledge, then, is the purpose or *telos* of inquiry. (This minimal conception of knowledge is, I believe, in keeping with the traditional use of the concept of knowledge in, for instance, Plato, Descartes, and Hume.)

I say that inquiry comprises procedures that have as their goal the generation of belief in *particular propositions,* because there are wider-ranging epistemic goals than knowledge: for example, rationality and wisdom. But such concepts take beliefs in groups or even consider the entire structure of beliefs of an epistemic agent. It can be rational or wise, as part of such an overall structure, to believe that *p*. But one cannot have the rationality or wisdom that *p*, as one can have the knowledge that *p*. At any rate, we might think of a theory of epistemic justification as doing two things: (a) giving general procedures for inquiry, and (b) setting out standards for evaluating the productions of inquiry, that is, particular propositional beliefs. A belief, then, is justified if it is produced (or, alternately, could have been produced) by a correct procedure for inquiry,[3] and to say that a belief is justified is to evaluate it positively along epistemic lines. I regard both projects as worthwhile. I am by no means arguing that an account of justification is unnecessary or unimportant; my claim, rather, is that such an account is not part of the theory of knowledge.

To continue with terminological groundwork, it is possible to distinguish two sorts of epistemological projects: normative and descriptive. Descriptive epistemology is essentially a branch of psychology or cognitive science. It asks: How *do* we come to believe or to know? What cognitive mechanisms and causal circumstances operate in the acquisition of belief or knowledge? What, for example, is the function with regard to belief and other propositional attitudes of perception, memory, and so forth? Normative epistemology, on the other hand, is what has usually been thought of as a distinctively philosophical discipline. It asks: How *should* we generate beliefs with regard to achieving knowledge? How should inquiry be conducted and evaluated? And what, exactly, is its goal, i.e., what is knowledge?

Now, these two sorts of investigations are by no means unconnected.[4] For example, it would be ludicrous to recommend (normatively) a procedure for generating beliefs which was literally impossible (descriptively) to apply. Thus, the normative project might be thought of as giving guidelines for the correct or optimal application of the procedures we do use to generate beliefs. I shall be concerned here with normative epistemology.

---

[3] On internalist views, that agent must be aware that the belief was so produced; on externalist views, the agent need not be aware of this fact.

[4] In fact, I agree with Alvin Goldman that they are very intimately connected indeed. See *Epistemology and Cognition* (Cambridge: Harvard, 1986).

II DEONTOLOGICAL AND TELEOLOGICAL CONCEPTIONS

The most elaborately developed normative theories are in ethics, and thus normative epistemology often relies on a parallel to ethics. Ethical systems have been divided into two kinds: deontological and teleological.[5] Proponents of the former think of moral action as what is done in obedience to principles that serve in turn no end that could be looked on as an overall moral goal. Obligation and permission are central concepts here. If I do only what is permissible (possibly, if I do it because it is permissible), or what is demanded by duty (possibly, if I do it because duty demands it), then I am not subject to ethical disapprobation even if the result of my action is disastrous. According to proponents of teleological ethics, on the other hand, an action is morally good when it conduces to some goal, for example, the greatest happiness of the greatest number, or if it is in accordance with some rule the observance of which so conduces. Similarly, there might be two sorts of normative epistemology: one that prescribes duties and permissions in generating beliefs (and other propositional attitudes) without regard to any overarching epistemic goal, and one that prescribes some goal for epistemic activity and recognizes the legitimacy of any procedure that conduces to that goal, or, alternately, of any procedure that accords with certain rules, the observance of which in turn conduces to that goal. I have asserted that knowledge is the goal of inquiry. But this supposes that inquiry *has* some goal, which would be denied by a proponent of deontological normative epistemology. So we had better start with a discussion of whether that position is plausible.

The taxonomy of normative epistemology suggested by this particular parallel to ethics has been developed by William Alston.[6] Because my discussion follows his to some extent, I should pause here to differentiate my use of terms from his. Alston uses the term 'deontological' to distinguish systems which epistemically prescribe, proscribe, or permit certain beliefs or belief-generating procedures, from systems he terms 'evaluative' which merely assess certain beliefs and procedures from the standpoint of some standard. He points out that it is not the case that all standards of evaluation depend on such concepts as obligation and permission, that not all standards carry with them the implication that the subject is praiseworthy for meeting them or blameworthy for violating them. For example, to say that some person is beautiful is to evaluate her appearance positively, but it is not to say that she is praiseworthy for her appearance, since she may not be responsible for it; it may be a genetic endowment (*ibid.*, p. 96). The relevant point here is that both sorts of systems (Alston's 'deontological' and 'evaluative') are what I term 'teleological'; he describes both as being directed to the goal of generating true belief and avoiding false belief.

---

[5] My characterizations of these ethical positions are sketchy and inadequate to many actual positions. The present point, however, is simply the parallel to epistemology, so we need not be concerned with capturing the subtleties of full-fledged ethical theories.

[6] "Concepts of Epistemic Justification," in *Epistemic Justification* (Ithaca: Cornell, 1989), pp. 81–114.

It may be a question, then, whether any philosopher has seriously held a deontological position in my sense, has seriously held that we have some epistemic obligations, but that there is no overarching goal of inquiry. Some extreme idealists and positivists, who identify truth and justification, may harbor such a view. If one has a coherence theory of truth and also a coherence theory of justification, for example, then one may simply count as knowledge whatever beliefs are generated by whatever procedures turn out to embody justification; if it was supposed to be a sheer fact that we ought to follow such procedures, if there were no further goal in mind, this would be a deontological position in my sense. The notion of knowledge is in some sense superfluous on this position; at least, it does not describe a distinctive purpose for inquiry above the fulfillment of certain duties or obedience to certain rules. Clearer examples of deontological views could be proposed: for example, believe all and only the propositions contained in the Bible, or in the writings of Mao.

Deontological views in my sense have, these days, few proponents, and seem on the face of it extreme and implausible. Their implausibility can be brought out in the following way. What is the *source* of our epistemic obligations? Or, to put it another way, is there any good reason to think that we have any distinctively epistemic obligations at all, in the absence of some overarching purpose for inquiry? The same problem arises for deontological moral theories, but here there are plausible, or at least fairly widely proffered, answers: our moral obligations derive from God, for example, or from the state. Again, it is possible that the very same sources yield our epistemic obligations. But to establish this, we would have to give good reasons to think that God does impose epistemic obligations, or to give an account of the "epistemic legitimacy" of the state. Furthermore, there no longer appears to be any distinction between moral and epistemic constraints on the generation of beliefs. There no longer appears to be any distinctively epistemological enterprise.

I shall take it, until such objections can be answered, that a purely deontological conception of epistemic norms is highly implausible. Indeed, no contemporary philosopher with whose work I am acquainted holds such a view. It might be asked, then, why I have bothered to discuss it at all. The point is just this: inquiry *does* have a purpose; only teleological conceptions of normative epistemology are plausible. But what about deontological conceptions in *Alston's* sense? Such conceptions formulate obligations and permissions, and award praise and blame, with regard to the way these obligations and permissions conduce to an overarching epistemic goal. I shall call such views, in keeping with my own rather than Alston's use of terms, ***injunctive teleological*** accounts of normative epistemology. They might be thought of by analogy to rule utilitarianism, where one's obligations are to follow rules, the observance of which tends to the greatest happiness of the greatest number. Rule utilitarianism has the advantage over act utilitarianism that it can yield relatively detailed injunctions on conduct, rather than the huge amorphous injunction to act so as to create the greatest good of the greatest number, which might be extremely difficult to apply in individual cases. Likewise, injunctive teleological views yield duties or permissions which, if observed, will conduce to knowledge, rather than the huge amorphous duty to generate beliefs so as to gain knowledge.

Alston attacks injunctive teleological accounts as presupposing what has been called "doxastic voluntarism," the view that our beliefs are, in one sense or another, under our voluntary control. After all, I cannot have an obligation to do something or to refrain from doing it if it is not up to me whether I do it or not. Injunctive teleological conceptions of justification indeed suppose doxastic voluntarism. And Alston's objections to doxastic voluntarism appear to me devastating. For a detailed discussion of the issue, I refer the reader to Alston.[7] Here, I recount a single example: when I seem to see a tree before me, and have no reason to think that, for example, I am dreaming, or that the perceptual situation is grossly abnormal, or that someone is replacing trees in my neighborhood with plastic replicas, it is simply not up to me whether I believe there is a tree before me; like it or not, I *shall* believe it. So I should hardly be praised or blamed for doing so. It should be noted that this criticism applies not only to injunctive teleological accounts but also to deontological accounts in my sense.

---

[7] See esp. "The Deontological Conception of Epistemic Justification," in *Epistemic Justification*, sect. II–VI

If this criticism is plausible, we are left with a conception of justification that Alston calls an "evaluative" or "strong position" account. These terms describe justification from the two angles mentioned above; to say that it is desirable to get into a strong position is to give a view about (a) the correct procedures for inquiry, while to evaluate someone's epistemic position positively (though not in deontological terms) is to apply (b) some account of evaluation with regard to these procedures. We ought not, on Alston's view, to alot praise and blame in epistemic matters, but some epistemic situations are better than others. (It is better to be beautiful than ugly, but, again, one is not usually praiseworthy or blameworthy in this regard.) On Alston's view, the most plausible way to formulate this conception is that one's belief is justified if it is based on adequate grounds, and having adequate grounds is a matter of being in a strong position to generate a true belief on the matter. It is a good thing to be in such a position, but one is not, or at least one is not always, responsible for getting into such a position. Now, it is not part of my current project to defend this conception of justification (though I regard it as plausible); rather, I want to use it in a certain way to elucidate what, on it, would be a plausible view of our epistemic goal with regard to particular propositions, that is, what would be a plausible theory of propositional knowledge.

III. TRUTH CONDUCIVENESS

It is widely held that our epistemic goal is achieving true beliefs and avoiding false ones about propositions with which we are epistemically concerned. (We have seen that Alston, for one, endorses that view.) That is, it is widely admitted that, on any good account of justification, there must be reason to think that the beliefs justified on the account are likely to be true. Indeed, proponents of all the major conceptions of justification hold this position. For example, the foundationalist Paul Moser[8] writes:

> [E]pistemic justification is essentially related to the so-called cognitive goal of truth, insofar as an individual belief is epistemically justified only if it is appropriately directed toward the goal of truth. More specif-

---

[8] *Empirical Justification* (Dordrecht: Reidel, 1985).

> ically, on the present conception, one is epistemically justified in believing a proposition only if one has good reason to believe it is true (*ibid.*, p. 4).

The reliabilist Alvin Goldman claims, similarly, that a condition on an account of justification is that beliefs justified on the account be likely to be true; he says that a plausible conception of justification will be "truth-linked" (*op. cit.*, pp. 116–21). And the coherentist Laurence BonJour[9] puts it even more strongly:

> If epistemic justification were not conducive to truth in this way, if finding epistemically justified beliefs did not substantially increase the likelihood of finding true ones, epistemic justification would be irrelevant to our main cognitive goal and of dubious worth. It is only if we have some reason to think that epistemic justification constitutes a path to truth that we as cognitive human beings have any motive for preferring epistemically justified beliefs to epistemically unjustified ones. Epistemic justification is therefore in the final analysis only an instrumental value, not an intrinsic one (*ibid.*, p. 8).

In fact, it is often enough taken to be the distinguishing mark of the fact that we *are* epistemically concerned with a proposition that we are concerned with its truth or falsity. That is what, on the view of many philosophers, distinguishes epistemic from moral or prudential constraints on belief, what distinguishes inquiry from other belief-generating procedures.[10] I have argued that a plausible normative epistemology will be teleological. And I have claimed that the conception that accounts of knowledge are attempting to analyze or describe is that of the epistemic *telos* with regard to particular propositions. It would follow that, if a philosopher holds that the epistemic *telos* is merely true belief, that philosopher implicitly commits himself, his own asservations to the contrary, to the view that knowledge is merely true belief.

I think that this is the case. That is, I think that, in the above passages, these philosophers *have* committed themselves implicitly to the view that knowledge is merely true belief, and that justification is a *criterion* rather than a logically necessary condition of knowledge. By a criterion, I mean a test for whether some item has some property that is not itself a logically necessary condition of that item having that property. Justification on the present view is, first of all, a means by which we achieve knowledge, that is, by which we arrive at true beliefs, and, second, it provides a test of whether someone has knowledge, that is, whether her beliefs are true. So, again, the present view does not make accounts of justification trivial, or unconnected with the assessment of claims to know. If our epistemic goal with regard to particular propositions is true belief, then justification (a) gives procedures by which true beliefs are obtained, and (b) gives standards for evaluating the products of such procedures with regard to that goal. From the point of view of (a), justification prescribes techniques by which knowledge is gained. From the point of view of (b), it gives a criterion for knowledge. But in neither case does it describe a logically necessary condition for knowledge.

---

[9] *The Structure of Empirical Knowledge* (Cambridge: Harvard, 1985).

[10] I hope to argue elsewhere that there are no non-epistemic belief-generating procedures in the relevant sense, that believing *p* and taking *p* to be true are the same thing. If this case can be made, then I think it can be shown to follow that knowledge is the *telos* of inquiry. So, though I take the latter claim here as an unargued premise, I think it can be decisively established.

Another way of putting the matter is this. If we describe justification as of merely instrumental value with regard to arriving at truth, as BonJour does explicitly, we can no longer maintain both that knowledge is the *telos* of inquiry and that justification is a necessary condition of knowledge. It is incoherent to build a specification of something regarded *merely* as a means of achieving some goal into the description of the goal itself; in such circumstances, the goal can be described independently of the means. So, if justification is demanded because it is instrumental to true belief, it cannot also be maintained that knowledge is justified true belief.

I shall now certainly be accused of begging the question by assuming that knowledge *is* the goal of inquiry. There is justice in this claim in that I have not gone very far toward establishing the point.[11] But I would ask my accusers at this point whether they can do better in describing the conception that theories of knowledge set out to analyze or describe without begging the question in favor of some such theory. And I ask also, if knowledge is not the overarching epistemic *telos* with regard to particular propositions, why such tremendous emphasis has been placed on the theory of knowledge in the history of philosophy, and just what function that notion serves within that history. If knowledge is not the overarching purpose of inquiry, then why is the notion important, and why should we continue to be concerned in normative epistemology above all with what knowledge is and how it can be achieved? If we want to withhold the term 'knowledge' from mere true belief, but also want to hold that mere true belief is the purpose of inquiry, then, first of all, I think that 'knowledge' will now merely be a technical term with a stipulated definition. And, second, I do not think it will be central to epistemology, since it no longer represents our epistemic goal. And, third, I think the stipulated definition will either be redundant (if justification is held to be truth conducive) or, as I shall argue, incoherent (if it is not).

---

[11] To repeat, however, I hope to establish this claim on completely independent grounds.

Now, it may well be held that justification is of more than instrumental value, because if we are not justified in believing *p*, though *p* is true and we in fact believe it, we may have false beliefs that lead us to *p*, and we may continue to generate false beliefs in the future. All of this is true, but it is irrelevant to the present point. Recall that I have characterized knowledge as our epistemic goal *with regard to particular propositions*. Insofar as *p* is concerned, this goal has been realized if *p* is true and we believe it. Insofar as we also have such goals as continuing to generate true beliefs, rendering our system of beliefs coherent, and so forth, it is desirable to have justified beliefs. But with regard to any particular proposition, our goal has been reached if we believe that proposition and it is true.

But I do not want simply to let the matter rest on a supposed agreement among some contemporary epistemologists that our epistemic goal with regard to particular propositions is true belief. Such epistemologists are agreed that knowledge is at least justified true belief. I think that Alston is right to think that the only plausible way to construe this claim is that knowledge is at least true belief based on adequate grounds, or true belief reached from a strong position. So perhaps the figures in question, on reflection, would describe the epistemic *telos* not as true belief but as true belief based on adequate grounds, or true belief reached from a strong position.

It must now be asked, however, *why* do we want to have adequate grounds? *Why* do we want to be in a strong position? This question ought to be misguided if true belief based on adequate grounds or true belief reached from a strong position is in fact the purpose of inquiry. For there is no good answer to the question of why we desire our ultimate ends. But the question is hardly misguided. In fact, we cannot even specify what it is to have adequate grounds except that these grounds tend to establish that the proposition in question is true; we cannot even specify what it is to be in a strong position except as being in a strong position to get the truth. This indicates that the purpose of inquiry can be formulated without reference to the notions of ground or position. Thus, on the views in question, believing the truth is in fact our overarching epistemic *telos* with regard to particular propositions, on the only plausible conception of justification. Hence, on these views, knowledge is merely true belief.

IV LYCAN'S EXPLANATIONISM

The only fully developed, contemporary account of justification of which I am aware that appears to steer clear of this consequence is that expressed by William Lycan in *Judgement and Justification.*[12] He proceeds by denying that truth is the only or the overall purpose of inquiry. His claim is that beliefs are justified in virtue of their explanatory role; they are justified by their coherence with previously held beliefs, by their simplicity, their scope in an explanatory framework, and so forth. And Lycan even argues that all inference relies on a (perhaps implicit) appeal to best explanation. For example, even in a simple *modus ponens* deduction, we are not forced to accept the conclusion; we always have the option of rejecting one of the premises. So Lycan concludes that all reasoning involves comparisons of plausibility. He describes justified beliefs as those which increase the overall explanatory coherence of one's system of beliefs.

Lycan's position is in some respects similar to BonJour's; he holds that we have "spontaneous," that is, noninferred beliefs that have some degree of prima facie justification (due to the explanatory virtue of conservatism, which states that we should not without some explanatory motivation change the beliefs we already do have). But Lycan's position differs in two crucial respects from BonJour's. First, Lycan goes further toward elucidating the notion of coherence in his description of the explanatory virtues. Second, whereas BonJour holds that any coherence theorist must give a "metajustification" of his view, that is, he must give reasons to think that beliefs that are justified on the view are likely to be true, Lycan demands no such thing. In fact, Lycan goes so far as to say:

> [I]f . . . appeal to simplicity [e.g.] is a fundamental epistemic method, then it is fundamental; no further question can arise regarding its "connection to truth." . . . For that matter, I do not in the first place accept the a priori assumption that epistemic justification must be a matter of quantitative relation to truth (*ibid.*, p. 148).

Lycan does indeed give a "metajustification" of his account, however. Simplicity, power, and the rest are useful on the evolutionary scale. They are adaptive. I am not sure how plausible this is, but the point is that, if we are to hold that justification is a component of knowledge, we need to take Lycan's general approach. That is, we must describe justification as conducing to some goal other than truth, or else lay down justification as a deontological desideratum. We cannot, as we have seen, coherently maintain that knowledge is the *telos* of inquiry, that justification is logically required for knowledge, and that accounts of justification must in turn be metajustified by their conduciveness to truth. And Lycan manages to avoid also the dangers of a deontological epistemology; he does not in fact regard simplicity or power as sheer desiderata, but as useful adaptations. Nor does his view assume doxastic voluntarism.

---

[12] New York: Cambridge, 1988.

Nevertheless, the view has severe drawbacks. To begin with, it makes the distinction between epistemic and other values obscure. In fact, it absorbs epistemic value into sheer utility, and it is not at all clear that this is a plausible line. It is true, as Lycan admits, that there is no a priori reason to think that a simple hypothesis is more likely to be true than an extremely complex one. But a simple hypothesis has less cognitive cost; it is efficient as an explanation. That may make it seem that it is better in the long run to believe simple hypotheses. But it is not clear that it *is* in fact, better, epistemically speaking, to believe simple hypotheses.

To get clear on the point, let us suppose (as seems likely, though Lycan is reticent here) that Lycan holds justification in his sense to be a necessary condition of knowledge, and that he holds truth also to be a necessary condition.[13] Then knowledge turns out to be (at least) true belief that is generated by adaptive explanatory techniques. But this seems odd; now that we recognize two primitive epistemic values, they may well conflict. For example, is it good to believe, in some circumstances, highly explanatory falsehoods? The account surely leaves some such cases strictly undecidable, since it describes both elegance and truth as intrinsic values. But is this plausible? Surely, we might want to say, though it can be useful to believe all sorts of falsehoods, it is always epistemically good to believe the truth. It may be useful, for example, for me to have a cognitive technique that causes me to believe that I have all sorts of positive qualities to an extremely high degree; I may be happier and more efficient if I believe myself to be extraordinarily intelligent, good-looking, humble, and so forth. But it is not a good thing epistemically to believe such things if they are false. (Fortunately, however, in my case they are all true.) In short, we seem to have a notion of epistemic value that is radically divorced from utility.

---

[13] It is only on this assumption that Lycan's position constitutes an alternative to the views described in the previous section. So I shall talk as though this is Lycan's position, though, again, he does not explicitly endorse it.

This raises the fundamental difficulty with Lycan's strategy; it makes prudential justification a component of knowledge. But if, as I say, we have a notion of epistemic justification in distinction from prudential justification, and if knowledge is essentially an epistemic matter, then Lycan's account leaves us confused. Either we ought to believe what is produced by procedures that are adaptive or else we ought to believe what is produced by procedures that are likely to yield true beliefs. But we cannot coherently demand that we follow both of these as ultimate aims, because they may and in fact will conflict. Then we are left with an internally incoherent concept of knowledge. Of course, we can subsume one to the other; we can assert that the truth is adaptive and that is why it should be believed, or that we should believe what works on explanatory grounds because it is likely to be true. But then, first of all, we need an account of why this should be so. And, second, we should now characterize knowledge univocally as belief in what is true, or belief in what is adaptive. If we take the second road, then I assert that we are going to hold all sorts of falsehoods to be known, and that is prima facie implausible.

In fact, Lycan keeps coming back to the suggestion that explanatory virtues *are* truth-conducive, although he often enough admits that there is no reason why this should be the case. This is because, I think, he shares the deep-seated intuition that knowledge and truth are bound up, that knowledge is a distinctively epistemic notion. He writes:

> Nor, please note, do I understand "rational" here as "likely to yield truth," for this would be . . . to introduce "true" at the beginning of our epistemological quest rather than see it fall out at the end. . . . [I]ntroducing "true" at this point would both threaten circularity or regress again and invite questions of the form "But what reasons have you to think that *U* leads to truth?" where *U* is a posited ultimate principle of theory choice. . . In the beginning, "rational" is a primitive term used to evaluate epistemic acts; particular principles are later seen to "tend toward truth," because the beliefs they produce are rational—not the other way around (*op. cit.*, pp. 155–6).

This passage strikes me as wanting to have it both ways. It seems that we simply ought to believe what is rational (simple, powerful, etc.), and then, as a sort of added bonus, as a matter of fact what we believe will happen to be true. But this is odd if, as Lycan admits, there is no reason to think that a simple explanation is more likely to be true than a very complex one. Now, as a matter of fact, the notion of truth itself here may well be disintegrating; Lycan's pragmatist tendencies might be extending toward taking truth to be warranted assertibility. The view would then face familiar and, I think, devastating problems, but I shall not address those here.[14]

But there is a further argument that Lycan gives for his position. He argues that we cannot be enjoined from doing what we cannot but do, and that we cannot fail to use explanatory criteria in generating beliefs:

> I suspect that it is biologically necessary to a sentient organism's survival that the organism organize its experience in the simplest, most coherent and expedient, and yet most powerful and comprehensive way that it can. . . . It thus comes to be lawlike that a viable organism operates, epistemically, according to the elegance principles . . . or something like them. . . . If the suspicion I have expressed is well-founded, the point is this: A normal human being cannot keep from forming beliefs in ways that have survival value, or from making use of the elegance principles in doing so—at least up to a point. And if that is so, then no one is obligated in any way to refrain from following those primitive procedures (*op. cit.*, pp. 126–7).

I would like to make several remarks about this passage. First of all, there is no reason whatever to believe that it is true that we are under adaptive necessity of adopting the most elegant explanation we can come up with, which is exactly what Lycan implies. In fact, that is a bizarre assertion; it claims quite generally that people never, or at any rate very rarely, adopt a less elegant hypothesis when a more elegant one is available, which is obviously false. Aside from the questions about the nature of simplicity (which Lycan leaves more or less completely obscure, and which I suspect cannot be clarified), I can surely resolve to adopt whatever hypothesis seems most likely to be true, however inelegant it may seem to be. It might be very elegant indeed, for example, to believe that the extinction of the dinosaurs resulted from a nuclear test on this planet by aliens, but a nice messy hypothesis involving climactic changes and disease might seem more likely to be true.

---

[14] All that would be necessary to refute this view is the description of a single case in which some agent is warranted in asserting *p* and *p* is false. Such cases are not hard to devise.

In fact, Lycan here goes further in his doxastic determinism than does even a rabid determinist, such as Alston. For Alston admits that we can do our best to apply conscientious techniques of belief formation. He simply denies that the fact that this is to some extent under our control is enough to render the beliefs that such techniques generate voluntary and hence subject to praise and blame. But Lycan asserts that we simply must use the techniques of belief formation that we do use. But this is false. Advances in methodology have been made, and can be voluntarily adopted. Nor do they all yield increases in elegance.

V CONCLUSION

Finally, I would like to formulate the argument that knowledge is merely true belief as a dilemma. If accounts of justification themselves must be metajustified by an argument to the effect that beliefs justified on the account are likely to be true, then justification is of instrumental rather than intrinsic value. That is, we want justification with regard to particular propositions not because of its intrinsic worth, but because it is a means to what we regard as having intrinsic worth, namely, true belief. If knowledge is the epistemic *telos* with regard to particular propositions, then it follows from such a position that knowledge is merely true belief. On the other hand, if justification is valued not for its truth conduciveness, but for its conduciveness to some other goal, for example, successful adaptation, or for that matter, if justification is itself proposed as an intrinsic goal (a demand of reason, for example), then knowledge is an incoherent notion. It gives us two goals for inquiry, which cannot always be realized simultaneously. As far as I can see, these two possibilities exhaust the possible accounts of knowledge as (at least) justified true belief. Again, either justification is instrumental to truth or it is not. If it is, then knowledge is merely true belief. If it is not, there is no longer a coherent concept of knowledge. Thus knowledge is merely true belief. Q.E.D.

CRISPIN SARTWELL

Vanderbilt University

A SENTENTIAL THEORY OF PROPOSITIONAL ATTITUDES*

I would like first to provide a general framework for the semantics of propositional attitude sentences. I shall then try to show how a sentential theory can very well be integrated within such a framework. I submit that the framework should take the form of a semantics that purports to give the assertability conditions of sentences. This entails that we should replace truth conditional semantics by a semantical theory that stipulates the conditions under which a sentence can be used and explains in this way "what role these sentences can play in our lives," to use Saul Kripke's[1] happy phrase. This general framework, one will have guessed, is the one that Kripke ascribes to the Ludwig Wittgenstein of the *Philosophical Investigations.*[2]

I would like first to make some general remarks on the semantics of assertability conditions. Such a semantics must take at least two factors into account: illocutionary force, which concerns the particular force with which the sentence can be uttered; and locutionary meaning, which concerns the conditions under which the sentences can be held true. These locutionary conditions can be identified with objective conditions that must prevail if the sentence is true or with verification procedures that serve as justification for the assertion.

I shall not deal with the problems related to the illocutionary conditions of use in this paper. Let me just say that propositional attitude sentences can give rise to the performance of a very wide variety of illocutionary acts. We can utter them intending to perform directly assertives, directives, commissives, expressives, and declaratives as long as certain propositional conditions prevail. First-person psychological sentences are particularly revealing in this respect. The five fundamental categories of illocutionary forces can be directly performed over them even if they are in the indicative mood. I mention this point just to indicate how illocutionary force comes into play within such a general framework. I allow illocutionary force to become an important ingredient in semantics without

---

* This paper was made possible partly by a Canada Research Fellowship coming from the Social Sciences and Humanities Research Council of Canada. I also wish to thank Tyler Burge, Charles Daniels, Kit Fine, and David Kaplan for their comments and criticisms.

[1] *Wittgenstein on Rules and Private Language* (Cambridge: Harvard, 1982).

[2] G. E. M. Anscombe, trans. (New York: Blackwell, 1953).

0022-362X/92/8904/181–201

suggesting that the indicative mood is a syntactic mark of illocutionary force. There are syntactic marks of illocutionary forces only for nondeclarative sentences, such as those in the imperative or interrogative moods. But the ones that are of interest for my study are all in the indicative mood, and I am taking for granted the fact that the indicative mood is neutral as to which illocutionary force can directly and literally be performed with it.

So the incorporation of illocutionary force in a semantics of assertability conditions does not come from the hypothesis that illocutionary force comes into play in the very syntax of indicative sentences, but rather from the fact that I have a broader perspective concerning what is to count as belonging to the field of semantics. Without the suggestion that they are part of the very syntactic features of each grammatical mood, illocutionary force can be part of my semantical account as long as they play a functional role that is not directly dependent upon the meaning intentions of speakers.

I also want to stress the fact that the locutionary conditions I alluded to do not involve any particular preference concerning their identification with truth conditions or verification procedures. As much as possible within the limits of this paper, I want to avoid having to make a choice between those two approaches. I said that the semantical approach I favor is one that spells out the assertability conditions of sentences, but this does not mean that the sentences in question do not have objective truth conditions. I just mean to suggest that truth conditions no longer constitute an essential ingredient in meaning and that having meaning for any particular sentence is not determined by the fact of having objective truth conditions. A semantics of assertability conditions could only be identified with a truth conditional semantics if we restricted the possible illocutionary forces to the assertive and if objective truth conditions were the only locutionary ingredient involved. But I precisely want to avoid having to make such an identification. I allow for the possibility of there being many different illocutionary acts performed with the use of those sentences, and for the possibility that locutionary conditions be identified with verification procedures in certain cases at least.

I

As I already pointed out, I intend to concentrate on the locutionary conditions of use of propositional attitude sentences. A sentence such as

(i) John believes that neutrinos have mass.

can either be used to describe a particular atomic belief or to assert the (general) existence of a belief that would satisfy certain descriptive features. According to the first reading of (i), it is asserted that John is in a certain belief relation with the sentence 'Neutrinos have mass'; according to the second reading, a certain sentence is said to be believed by John and can be translated as 'Neutrinos have mass'. This ambiguity between the two readings is semantical and affects the 'that' in (i). The presence of a that-clause in any given statement always produces a nominalization of the subordinate sentence and forms a complex name for a sentence, but this can take place in two radically different ways. When the whole sentence is used to report a particular atomic belief, the that-clause behaves a bit like a complex demonstrative expression or like the occurrence of a sentence within quotation marks. We could choose to describe it as a demonstrative expression or as a quotation, depending on whether we choose to describe belief as a relation toward an utterance or a sentence type. For my own part, I prefer the latter and shall therefore describe this first use as reporting a belief toward a particular sentence type, but I could also adopt the other approach. If I had done so, the first use would look very much like Donald Davidson's[3] paratactic theory of indirect speech.

The second use I want to introduce also involves a nominalization of the subordinate clause but it ultimately must be contextually eliminated in terms of a quantified statement. We proceed in this case as in Bertrand Russell's theory of descriptions and eliminate the expression 'that neutrinos have mass' as a pseudo proper name. This yields a general statement asserting that a certain sentence type is believed by John and is a translation of the sentence type 'Neutrinos have mass'.

The deep motivation for making such a distinction between atomic and general belief reports is that we want to allow for the possibility of reporting within the limits of our own language someone else's belief with his own understanding of linguistic meanings or his own principles of maximal rationality. This is what happens in the case of atomic belief reports when we simply want to describe the belief relation that an agent entertains with a particular sentence. It must be contrasted with the case where we assert the general existence of a belief and provide a translation for its content by exploiting the resources of our own conceptual scheme. The distinction between the two reports is thus founded upon the possibility of allowing for a relativity of conceptual schemes within a single language and, therefore, of making a distinction between language and theory. This is, of course, an antiholistic claim but I am prepared to make it.

---

[3] "On Saying That," in his *Inquiries into Truth and Interpretation* (New York: Oxford, 1984), pp. 93–108.

There are therefore two ways of representing (i). When we use the word 'that' as a name for the subordinate clause, we have

(ii) John believes 'Neutrinos have mass'.

and when we use it as a pseudo proper name, we get a general formula like

(iii) (Σp) [(believes (John, 'p')) $\wedge$ ('p' is translated as 'Neutrinos have mass')]

II

I shall return shortly to the motivations for choosing such a logical form. In particular, it will be instructive to give the motivation for introducing a substitutional rather than an objective quantifier in the analysis. For the moment, I want to mention a second ambiguity affecting this time the verb 'believes' in our original sentence. The problem is that, no matter what sentence is believed, we must distinguish between its character, or linguistic meaning, and its content, which is identical either with its truth conditions or verification procedures. The statement that counts as the object of belief can now be mentioned in order to specify either its character or its content. If we mention the sentence in order to specify its character, the whole belief sentence reports a belief relation between the agent and a certain linguistic meaning under a mode of presentation that is a certain verbal form. If we mention the subordinate clause in order to specify its content, the agent is then described as being in a relation with the content of the sentence mentioned and, in this case, the sentence no longer behaves as a mode of presentation. The ramification of sense into character and content is therefore responsible for an ambiguity affecting the attitudinal verb.[4] We can speak of two uses of those verbs. I shall speak of an intentional use of the verb in the case where the whole sentence relates an agent to a certain linguistic meaning via a sentence type, and speak of a material use of the verb when the whole sentence describes a relation between an agent and a certain content and the sentence used no longer serves the purpose of specifying any mode of presentation at all for the agent.

---

[4] Cf. Kaplan, "Demonstratives," in J. Almog, J. Perry, H. Wettstein, eds, *Themes on Kaplan* (New York: Oxford, 1989), pp. 481–563.

On the basis of these two uses, propositional attitude sentences can be used to assert many different things. For example, we shall be tempted to make an intentional use of the verb 'believes' in order to report the actual content of an intentional belief, while the material use of the verb will ideally serve the purpose of reporting a functional teleological state of the agent. In the first case, the ascription of belief will be founded either upon sincere assent or dispositions to assent sincerely to a sentence. We could also include in the class of intentional beliefs those which are trivial logical consequences of sincere assents or of dispositions to assent sincerely to a sentence. In the case of a material attribution, we try to represent a certain functional state of the agent. It is as though the agent were described as trying to adapt himself to certain characteristics of the environment, whether these characteristics exist or not. As an example of this second use of the verb, we could mention the case of a dog barking at the bottom of a tree functionally behaving as though there were a cat in the tree. Another example of what I call a material belief is provided by the case of unconscious beliefs that the agent is trying to repress, perhaps without succeeding very well. We could also mention beliefs that are nontrivial logical consequences of an agent's already existing set of beliefs.

It is very hard to define very clearly these two uses. The distinction serves to capture a difference between the true intentional beliefs of the agent and those which he "in fact" has, whether he likes it or not and whether he knows it or not. We could, however, mention three features that can serve together as a criterion for discriminating between the two uses. Two of those features have already been mentioned. First, the relatum is different depending on whether the use of the verb is intentional or material. In the first case, the relatum is a linguistic meaning under a certain verbal mode of presentation; in the second case, it is a certain content. Second, the nature of the belief relation is different whether we are dealing with intentional or material uses of the verb. In the case of an intentional use, we could talk about a relation of directly grasping or apprehending a certain intentional content; in the case of a material use, we are dealing with a purely functional relation that does not involve grasping or apprehending any content whatsoever. Third, the two kinds of beliefs differ also by the fact that only intentional beliefs are reflexive and entail a first-person authority. If an agent intentionally believes that *p*, he represents himself as someone who believes that *p* and, as matter of fact, he knows that he believes that *p*. This is not the case for the material use. In the case of a material belief, the agent need not even have a reflexive representation of his own belief.

Now, since we have a second ambiguity, it follows that (i) can serve to describe at least four different possible states of affairs; (ii) and (iii) can be used to report intentional or material beliefs. We have first two distinct atomic belief reports, namely,

(iv) John believes_i 'Neutrinos have mass'.

and

(v) John believes_m 'Neutrinos have mass'.

where the subscripts indicate intentional and material beliefs. There are also two distinct general belief reports, namely,

(vi) (Σp) [(believes_i (John, 'p')) $\wedge$ ('p' is translated as 'Neutrinos have mass')]

and

(vii) (Σp) [(believes_m (John, 'p')) $\wedge$ ('p' is translated as 'Neutrinos have mass')]

I would use (iv) if I were to report directly John's intentional belief and wanted to use John's own concepts of neutrinos and mass, i.e., the ones involved in his own linguistic group. He could have intended something different from an extremely light electrically neutral particle having a certain weight. And if, for instance, I had directly inferred from John's calculating behavior a functional belief to the effect that neutrinos have mass without presupposing my own principles of rationality, I could have used (v).

These reports are not meant necessarily to capture John's idiolect but rather to allow for the presence of different dialects or different principles of rationality within a single linguistic community. Saying that (i) might be used to convey such different linguistic meanings or contents suggests that different conceptual schemes can come into play within the confinement of a single home language. It suggests that I can use a verbal form from my own dialect and still be able to express a very different conceptual scheme. Good examples of this take place when an English speaker is able to master the philosophical vocabulary of other English speaking philosophers having views that are quite different from his own, or when she is a classicist but is able to make inferences within an intuitionistic logic.

Of course, these differences can only make sense if the speaker and agent share a minimum linguistic vocabulary, and if within the limits of what I have called "linguistic meaning" there is a core shared both by the speaker and the agent. Rejecting a holistic picture of language will enable me to allow for the possibility that two speakers sharing the same linguistic core could not only have a different truth theory but also a different ideology or different principles of rationality. In short, I want accept the relativity of conceptual schemes.

In a general intentional belief report, such as the one described in (vi), the speaker makes an implicit use of a translation manual and tries to provide in his own dialect a "good" translation that captures the linguistic meaning of the sentence actually believed by the agent. For instance, a French scientist could have assented to the sentence 'Les neutrinos ont une masse' and (vi) could be used to convey this information. In a general material belief report, such as the one described in (vii), he tries to provide a translation that captures in his own dialect the same content as the one that is materially believed by the agent. For instance, I could have heard of an Italian scientist who believed that the things he had just located in the laboratory under Mont-Blanc have mass, not knowing that these things are neutrinos. It would then be wrong to say that he intentionally believed that neutrinos have mass, but correct to ascribe a material belief to that effect.

In both cases, however, the speaker is making use of a principle of charity and is trying to maximize agreement with the agent. In such an intentional belief report, he is maximizing agreement concerning his understanding of linguistic meanings; in the case of a material report, he is maximizing agreement over content. In other words, he is either projecting onto the agent his own understanding of the linguistic meanings and presupposing that a good one-to-one translation of those linguistic meanings is possible between the two dialects, or projecting the same functional relations toward content and presupposing that a good one-to-one translation could be made that preserves the content of the original sentence.

III

I should consider one last distinction: between reports that are used to represent basic beliefs, and those which are used to represent derived beliefs. Accordingly, the atomic intentional basic beliefs are going to be those which are ascribed on the basis of sincere assent to sentences (or of dispositions to a sincere assent) that do not presuppose any inferential act on the part of the agent. The atomic intentional derived beliefs are those which are going to be inferred from atomic intentional basic beliefs of the agent. For instance, if the agent asserts or is in a disposition to assert '*p*' and asserts or is in a disposition to assert '*p* materially implies *q*', and if he does not infer these from any other previous beliefs, we have basic intentional atomic beliefs. But if the agent is in addition disposed to assent to '*q*' on the basis of those previous assertions or dispositions, we are then dealing with derived intentional basic beliefs. In other words, the atomic intentional derived beliefs are going to be those which presuppose basic intentional beliefs and a certain understanding of the linguistic meaning of expressions. It is important to point out, however, that this understanding is the one presupposed to be on the part of the agent and does not necessarily coincide with our understanding.

The same kind of remarks apply mutadis mutandis to atomic material beliefs. There are those which are ascribed directly from the behavior of the agent and those which are ascribed on the basis of previous ones. For instance, I can directly infer from the behavior of a dog barking at the bottom of a tree the material basic atomic belief that the thing seen before is now in the tree. On this basis, I can then get a derived material atomic belief to the effect that there is a cat in the tree. The derived atomic material beliefs will include those which we ascribe while presupposing a certain inferential capacity on the part of the agent. Again, the inferential capacity is not necessarily going to coincide with our own.

These differences between basic and derived beliefs also occur in the case where the belief reports take the form of general quantified statements. The difference is that our own understanding of linguistic meaning, in the case of an intentional report, and our own inferential capacity, in the case of a material report, will be the things that we are going to presuppose in order to generate derived beliefs from those which are basic.

I could multiply distinctions like these, but I am going to stop here. The three distinctions I have introduced will serve my purposes for the time being. I shall just mention, in passing, other distinctions that I chose to ignore in order to simplify my presentation. I earlier said I would not discuss the different varieties of illocutionary acts that can be performed with propositional attitude sentences. I shall also leave aside problems related to the fact that we are using a third-person rather than a first-person psychological sentence. We would have to deal ultimately also with the distinction between de dicto and de re beliefs and with sentences involving iterated belief operators.

Before ending this taxonomic presentation, I wish to draw your attention to another very important feature of our belief reports. I am thinking about the pragmatic presuppositions that come into play when we are making use of these sentences. For instance, when we are using a propositional attitude sentence to report an atomic intentional basic belief, we can presuppose that the agent has a complete understanding of the linguistic meaning of the sentence, but we could as well cancel this presupposition. It is not information contained in the literal meaning of the belief sentence. As I have described them, intentional attitudes entail a relationship with linguistic meanings that are mediated by language. A certain verbal form serves the purpose of a mode of presentation. This is perfectly compatible with the possibility that the agent has an incomplete understanding of the sentence that behaves as the object of his attitude. Therefore, when we are taking for granted the fact that the agent has a complete understanding of the linguistic meaning attached to the sentence by the members of his own linguistic community, we are making a pragmatic presupposition that can by definition be cancelled. When we are making such a pragmatic presupposition, however, our intentional reports will involve an assumption of minimal subjective rationality on the part of the agent whether or not the rationality principles are similar to our own.

By ***minimal rationality*** I mean a rationality that is based solely upon an understanding of the linguistic meanings attached to the logical connectives and quantifiers by the members of the community to which the agent belongs. Since I wish to promote a view of language that involves the idea of a general linguistic practice governed by indeterminate semantical rules attached to the primitive expressions, i.e., semantical rules that do not anticipate all the applications of the expressions, the rationality I am talking about is not holistic.

As far as basic atomic material beliefs are concerned, we do not need to make such presuppositions. The reason is that the agent is only in a functional relation with the sentence. An agent can be in such a functional relation to a sentence even without being able to speak a language. The sentence mentioned in the belief report merely serves to describe the specific teleology involved in his functional states. We can pragmatically presuppose, however, that the agent has a certain "know-how," i.e., a certain maximal inferential capacity. The degree of complexity of these maximal inferential capacities will vary depending on whether we are presupposing the know-how of a community of chimpanzees, human beings, or artificially intelligent devices.

When we are ascribing material beliefs, we are often presupposing a variety of degrees of subjective rationality on the part of the agent. But since this is also a pragmatic presupposition, it could be cancelled. The ascription of material atomic beliefs does not entail semantically a subjective rationality on the part of the agent. But if we were to make such a pragmatic presupposition, we could describe the beliefs of the agent in a fashion that would rather fit well with a holistic picture. As it was mentioned earlier, my view remains quite opposite to a holistic account since I allow for intentional and material beliefs quite independently from any maximal rationality assumption. But I do not wish to elaborate any further on my rejection of holism since this would lead us far beyond the scope of the present paper.

The two pragmatic presuppositions that I have described—the one concerning the knowledge that the speaker has of the linguistic meanings that are involved in our intentional reports, and the maximal rationality assumption involved in our material reports—often accompany our belief ascriptions. It is perhaps worth noting that it is only because we are making these presuppositions that we are able to allow for derived beliefs. It is only because we are presupposing a complete understanding of the linguistic meanings that we can in the context of an intentional ascription allow for the substitution of synonyms or draw an inference based upon the linguistic meanings attached to the logical connectives and, therefore, make derived intentional ascriptions. It is only by presupposing a certain maximal inferential ability on the part of the agent that we can in the context of material belief ascriptions make derived material ascriptions, such as those which are logical consequences of basic material beliefs.

Pragmatic presuppositions also come into play in our use of quantified statements that assert the general existence of intentional or material beliefs. I have chosen to represent this use by introducing a translation relation. This is because, in such a use, the subordinate clause is the result of a translation in the dialect of the speaker. When the speaker is using the sentence in this way and is attempting an intentional report, he very often presupposes his own understanding of the linguistic meaning of expressions in the subordinate clause. Most of the time, speakers presuppose that they have a complete understanding of the words they are using but, of course, we cannot exclude the possibility of an incomplete understanding. It is once again a pragmatic presupposition that he may want to cancel.

Once this pragmatic presupposition is made, however, the speaker may ascribe, if he wishes to do so, derived intentional beliefs that result, for instance, from the substitution of synonyms within his own dialect. He will very often want to substitute for the expression that is a good translation of the one involved in the sentence believed by the agent. In these instances, the general intentional belief ascriptions are those which the agent would entertain if he were a member of the same linguistic community, in the same context and the same perspective as the speaker's. If the agent had used the translation manual actually used by the speaker, the beliefs that are ascribed to him by the speaker would involve sentences that he would be disposed to accept.

But the speaker will sometimes also want to ascribe derived beliefs resulting from the substitution of expressions that, relative to the context and to the perspective of the speaker, determine the same content as the one that is materially entertained by the agent. And when we assert the existence of such and such a material belief in a sentence that can be translated in such and such a way, we very often presuppose as speakers a certain maximal inferential capacity. We presuppose that we have access to objective rational principles. As soon as we make such presuppositions, we are in a position to ascribe to the agent general material beliefs that are derived and not only basic. These new belief ascriptions are a consequence not only of a translation that is preserving the content of the original sentence in the dialect of the speaker, but also of projecting onto the agent the rationality principles the speaker is willing to accept.

IV

Many things could be said concerning the relationship between all those different locutionary conditions and the project of a sentential theory of propositional attitudes. But I shall raise only one main issue. I would like to claim that this very large variety of uses can be accounted for, not in spite of the fact that we are dealing with a sentential theory but rather precisely because we are using such a theory.

This claim might sound surprising in light of the difficulties with which traditional quotational theories are confronted. Many authors have argued for quotational theories,[5] but they have for a very long time now fallen in disrepute. For one thing, they very often find themselves inextricably tied with an empiricist epistemology or with a nominalistic ontology, which gives them an unnecessary air of controversy. The truth is that a sentential theory can be formulated while preserving a relative neutrality regarding these issues. The major difficulties are rather of a logico-semantic nature. Philosophers[6] have been concerned about the fact that semantic paradoxes could be creeping through the back door. It has also been maintained that quotational theories in general would fail a translation test.[7] Others have claimed that quotational theories will discourage any attempt at constructing an epistemic logic.[8] The charge is also very often made that, within the framework of a sentential theory, belief contents are too finely discriminated.[9] At the same time, the theory makes it impossible to discriminate between homophonic beliefs that differ in content, whether this is because the sentence believed is ambiguous or because it belongs to two different languages (*ibid.*). There is also the difficulty of accounting for the fact that the content of a belief is not merely a pure verbal form, but rather something that has certain semantic properties, without reintroducing things that look very much like propositions.[10] There is also, of course, the difficulty of accommodating de re attitudes within the framework.[11] On top of all of this, the theory is in any case confronted with even more obvious difficulties. It has an artificial character and is unable to explain why we ascribe propositional attitudes to creatures that are devoid of any language capacities.

---

[5] E.g., Wittgenstein, *op. cit.;* Rudolf Carnap, *Meaning and Necessity* (Chicago: University Press, 1947); Davidson, *op. cit.;* Israël Scheffler, "An Inscriptional Approach to Indirect Quotation," *Analysis,* x (1954). 83–90.

[6] E.g., Richard Montague and David Kaplan, "A Paradox Regained," in Richmond Thomason, ed., *Formal Philosophy: Selected Papers of Richard Montague* (New Haven: Yale, 1974), pp. 271–85; Thomason, "Indirect Discourse Is not Quotational," *The Monist,* LX (1977) 340–54.

[7] Cf. Alonzo Church, "On Carnap's Analysis of Statements of Assertion and Belief," *Analysis,* x (1950): 97–9

[8] E.g., George Bealer, *Quality and Concept* (New York: Oxford, 1982).

It is going to be useful to examine briefly the particular formal representation I wish to introduce for propositional attitude sentences. When a sentence such as (i) is used to assert the general "existence" of a belief that neutrinos have mass, it can be represented as

(viii) (Σp) [(believes (John, 'p')) $\wedge$ ('p' is translated as 'Neutrinos have mass')]

I leave aside the distinction between the material and intentional readings. This particular analysis reminds us a bit of the one that was once defended by Rudolf Carnap,[12] but many very important differences need to be mentioned. First, the quantifier is substitutional and not objectual. Second, the verb 'believes' must not be reduced to a disposition to assent to a sentence as prescribed by Carnap's empiricism. This is because we allow for a material reading as well as an intentional one. It is only for the latter that we may find a disposition to assent. Third, the substitutional variable "stands for" a sentence in use and is not an ambiguous name of sentences. The sentences are those which belong to the substitutional class and these semantically presuppose their semantical rules. Therefore, a substitutional formula cannot express anything meaningful unless its substitutional instances, i.e., the formulas that result from the substitution of the variable by an expression in the substitutional class, are themselves meaningful. But the substitutional instances are going to be meaningful only if the subtituenda themselves, i.e., the sentences in the substitutional class corresponding to the variable, are already meaningful. A constraint of meaningfulness is therefore already registered in the substitutional classes and guarantees the meaningfulness of the whole substitutional formula. Now, since the sentences that occur within quotes in a substitutional formula semantically presuppose their semantical rules, it is not necessary to mention explicitly the fact that the sentence belongs to a language, as it is the case for Carnap. Finally, the relation of intensional isomorphism is replaced by a weaker relation of translation. These apparently subtle differences have, I believe, enormous consequences.

---

[9] Cf. Max Cresswell, "Quotational Theories of Propositional Attitudes," *Journal of Philosophical Logic*, IX (1980): 17–40

[10] Cf Cresswell, *op. cit.;* Stephen Schiffer, *Remnants of Meaning* (Cambridge: MIT, 1987).

[11] Cf. W. V. Quine, "Quantifiers and Propositional Attitudes," this JOURNAL, LIII, 5 (March 1, 1956): 177–87.

[12] *Meaning and Necessity* (Chicago: University Press, 1947).

It is very important to notice that any expression can belong to a substitutional class. Nothing prevents the left-hand parenthesis or a comma from appearing among the members of any given class. But if the substitutional instances of a formula like (viii) are to be syntactically well-formed, the substitutional class must contain full sentences only. In addition, a moment of reflection reveals that, in order to express truth conditions or verification procedures, (viii) must also be evaluated relatively to a substitutional class in which the sentences are already meaningful. There are two cases to consider. If (viii) is used in the intentional sense, it can express a content only if the substituenda have at least a linguistic meaning. If (viii) is used in the material sense, it can express a content only if the substituenda themselves express either truth conditions or verification procedures.

So it depends what we mean by 'meaningful'. A substitutional formula not satisfying any such constraints can certainly be 'meaningful' if by that word we understand 'having a linguistic meaning'. But is it meaningful in the sense of expressing truth conditions (or verification procedures)? A sentence has truth conditions only if it shows the conditions under which it is true. Now, a formula like (viii) would do that if it would show that it is true or that it is false under each of its substitutional instances. As an existential statement, it is true only if the disjunction of its substitutional instances is true. (A universal substitutional quantification will be true only if the conjunction of substitutional instances is true.) But the disjunction will be true only if each of the disjuncts have truth conditions. If (viii) is an intentional report, the disjuncts will express truth conditions only if the quoted sentences have at least a linguistic meaning. If it is a material report, it will express truth conditions only if the quoted sentences in the substitutional instances themselves express truth conditions.

It is very important not to confuse the substitutional quantifier with an objectual quantifier over verbal forms and their meanings. In such an objective quantified formula, sentences would come "wearing their meanings on their sleeves." In a substitutional language, subtituenda semantically presuppose their meanings. These are therefore not asserted explicitly in the formula. Their existence only comes into play as a constraint that must be satisfied if the substitutional formula is to express a content. A sentence in use generally does presuppose semantically its semantical rules. So it is correct to say that substituenda do enter sentences with such semantic presuppositions. Once again, this is because their having meaning is a constraint on the very meaningfulness of the substitutional quantifier.[13]

These constraints of meaningfulness must be satisfied by the substituenda even when they occur within the context of quotation marks as in (viii). This brings us to another important feature of substitutional quantification and it is the status of quotation marks in such a language. This peculiar feature is present not only when a substitutional variable occurs within quotes, but also whenever a particular lexical item does. Atomic belief reports, such as in (ii), (iv), and (v), must be seen as substitutional instances of the formula '(Σp) (believes (John, 'p')', and so whatever I have to say about the behavior of quotes in a general formula also applies to quotes in particular substitutional instances. Now, whenever a sentence occurs within quotes in a substitutional instance, it is as though it were simultaneously used and mentioned. We have already seen that the substituenda enter the substitutional instances while presupposing their semantical rules. We must now see that they behave a bit like arguments for the quotation function. When a quotation occurs in a substitutional formula, the only way to encode it is first to understand the sentence as used, and then take it as an argument for the quotation function.

The choice to represent a propositional attitude sentence as a substitutional formula has the advantage of revealing the twofold nature of the sentences that are objects of belief, i.e., of being simultaneously used and mentioned. In contradistinction with an objectual representation in which the sentence is only mentioned, our own characterization circumvents some of the traditional difficulties. The sentences occurring within quotes presuppose their semantic interpretation and are therefore modes of presentations for them. As opposed to the Carnapian classical way of formulating the quotationalist view, the agent is not described as being in a relation with a pure verbal form relative to a language *L*, but rather with a linguistic meaning via a verbal form. If a change of vocabulary is required, I do not mind to call my theory a "quasiquotational theory." The expressions 'quotational' or 'sentential' still seem to me to be accurate renditions, for it was never in the mind of the classical quotationalist philosophers to suggest that we believe verbal forms without contents.

---

[13] Cf. M. Sainsbury, "Russell on Constructions and Fictions," *Theoria*, XLVI (1980): 19–36.

In (viii), the substitutional variable occurs in a quotational context and quotational contexts are very often described as the opaque contexts par excellence. Whether or not this is true, it is natural to allow substitutional quantification in these contexts. The result of putting a substitutional variable into quotes does not yield a name for the variable. It is rather more like a quotation function occurring all by itself. Quantifying substitutionally into a quotation context is simply binding the variable of the quotation function. Quotes are like descriptive functions that take objects as arguments and also have objects for values, unlike propositional functions whose values are truth values. They form a peculiar sort of descriptive function first, however, because the value of the function is in these cases precisely its argument. Another peculiar feature is that, contrary to the descriptive functions in *Principia Mathematica,* the quotation function is primitive and cannot be defined.

It is very important not to confuse the substitutional quantifier with an objectual quantifier over expressions. A formula like (viii) does not assert the existence of an expression. Existence is not asserted but presupposed semantically. The existence of expressions in the substitutional class is a condition that must be met for the formula to have any truth value whatsoever. The existence or nonexistence of expressions in the substitutional class determines whether the formula has a truth value or not. If an objectual quantifier were used, an empty domain would yield falsity. By contrast, a substitutional formula like (viii), evaluated relatively to an empty substitutional class, has a truth-value gap. The formula is neither true nor false but remains meaningful. This does not imply, however, that it is meaningful even in the absence of the concept of a language. The notion of a language understood in the sense of a system of syntactic and semantic rules is itself implied in the very notion of a substitutional class whether it is empty or not. Since attitudinal verbs are represented as ranging over substitutional classes, this means that we cannot make sense in belief reports of the notion of belief without the notion of a language.

Now, putting these two features together—the constraint of meaningfulness on the substituenda and the particular behavior of quotation marks—reveals yet another way to discriminate between the substitutional quantifier and an objective quantifier over expressions. One could want to make the suggestion that a substitutional quantifier is nothing more than an objective quantifier over expressions that semantically presuppose their semantical rules. According to this account, however, the believer is related to a pure verbal form and has no relation whatsoever with its meaning. In other words, we lose an important feature gained by the use of a substitutional representation and it is the fact that the quoted expressions are used and not only mentioned. By 'used', I do not mean, of course, that what is quoted is a concrete inscription. It is more like Peter Strawson's notion according to which a use is a sentence type in a context. The substituenda occurs in a quotation context as it would occur in any other linguistic context. For instance, no one will deny that the substituenda are "used" in substitutional instances, such as those for '(Σp) (John believes that p)'. The result of replacing the variable by a sentence like 'Neutrinos have mass' yields the sentence 'John believes that neutrinos have mass', and no one will deny the fact that in the latter the subordinate clause is used. My point is simply that, when they occur within the context of quotation marks, substituenda are used in the same way except that they are at the same time "arguments" for the quotation function.

This apparently subtle difference with an objective quantification over expressions is what justifies us in saying that, in a substitutional formula like (viii), the verbal form occurs as a mode of presentation for its meaning. Since the substituend is used and presupposes its semantical rules even when it appears in the context of a quotation function, we can give a coherent representation of the idea that the verbal form behaves as a mode of presentation and this is precisely something that an objective representation cannot do.

V

I shall now try to show how the account solves many of the traditional puzzles generally associated with a quotational approach. The theory enables us to account for the fact that we can ascribe beliefs to creatures that are unable to speak. Material ascriptions of beliefs, as we have seen, do not entail that the agent intentionally grasps any linguistic contents. Material ascriptions only serve the purpose of describing functional states of the agents and the linguistic component marks their specific teleology.

We are also in a position to solve another well-known difficulty. As previously mentioned, the substitutional variables range over classes containing expressions that presuppose their semantical rules. If these expressions are ambiguous or belong to different languages, they will presuppose different semantical rules. When the expression occurs in a given substitutional instance, the speaker will very often presuppose, in accordance with the literal meaning of the sentence, a particular lexical meaning. And this is how we are able to account for the differences between belief ascriptions made with a unique verbal form. The answer is again that the expression comes fully equipped with all its semantical rules and that we are presupposing only one. The fact that our ordinary belief ascriptions are sensitive to ambiguity does not tell anything against sentential theories. It simply reveals that speakers who are making belief ascriptions must make pragmatic presuppositions in order to achieve a specific claim. But this is not anything peculiar, since it concerns all fragments of natural languages and not only propositional attitude sentences.

Another classical objection loses its raison d'être. This is the claim that, within the framework of a quotational theory, belief contents are too finely discriminated. There seems to be no way to allow for the substitution of synonyms. But as we have already seen, this is no longer true concerning the present approach. Since the substituenda occur in substitutional instances with their associated semantical rules, the theory allows for uses made in accordance with the literal meaning in which it is presupposed that the speaker has access to the linguistic meaning of the sentence believed. The account I have just sketched indicates that it is only in the case of basic intentional atomic reports, in which the speaker does not presuppose that the agent has full understanding of the sentence believed, that a homophonic rendering of the belief content is required and prevents any substitutions to be made.

On the other hand, the theory provides at the same time neat solutions to philosophical puzzles, such as those discussed by Benson Mates, Tyler Burge, and Kripke[14] concerning the apparent occasional failure of interchangeability between synonyms or between expressions that are good translations of each other. As a matter of fact, traditional sentential theories have always retained a certain plausibility precisely because they were able to explain why sometimes even synonyms cannot be substituted with each other in belief contexts and why there seems to be no criterion of synonymy good enough to block such puzzles. We can explain why some intentional belief reports seem to resist interchangeability of synonyms no matter what the criterion is. In an intentional report, it is as though the verbal form were a mode of presentation for the linguistic meaning. I said that the substituenda occur with their associated semantical rules, but it must be emphasized that they do not "wear them on their sleeves" as they would if we had decided to quantify objectively over sentences and their associated meanings. Our account is not only compatible with the possibility of an incomplete understanding, it provides also an explanation for it. If the agent had direct access to concepts without the mediation of verbal forms, it would indeed be difficult to explain how someone could have an incomplete understanding of concepts.

---

[14] Mates, "Synonymity," in Leonard Linsky, ed., *Semantics and the Philosophy of Language* (Chicago: Illinois UP, 1952), pp. 111–36, Burge, "Belief and Synonymy," this JOURNAL, LXXV, 3 (March 1978): 119–38; Kripke, "A Puzzle about Belief," in Avishai Margalit, ed., *Meaning and Use* (Dordrecht Reidel, 1979), pp. 239–83.

Although I cannot show it in detail in the present paper, it is clear that the theory will enable us successfully to pass Alonzo Church's translation test. As it was said above, there is no longer a need to make an explicit reference to the languages to which the quoted sentences belong. Quotations are not, under the present account, names of pure verbal forms (nor of verbal forms plus linguistic meaning), but rather sentences in use as substituenda within the scope of a quotation function. Good translations will therefore not require a homophonic translation of the quoted sentences. We are thus in a position to maintain a sentential account without having to face difficulties, such as those which Church saw in Carnap's proposal. There are no longer any problems involved in allowing for the same transformations to take place in the quoted sentence of the definiens and in the subordinate clause of the definiendum. In other words, the definiens can be translated in another language without losing its status of definiens. Its translation in another language can express the same thing as what is expressed by the translation of the definiendum in the same language.

Another traditional objection concerns the possibility of developing an epistemic logic. According to the objector, this is not an option available to the quotationalist. The objection is that, in a sentential account, no inferential patterns can be allowed. But it now appears to be a wrong diagnosis. Within the framework of the present theory, it is even a bit anachronic. There is indeed a difficulty under the present account to provide a univocal answer to the question concerning the validity of the rules of introduction and elimination for the logical connectives in belief contexts. The reason is not, however, that the sentential account does not permit any inference to take place, but rather that there cannot be an absolute answer to the question. It all depends on the use being made of the sentence. Are we using the sentence to report an atomic belief or to assert the general existence of a belief under a description? Are we making an intentional or a material use of the psychological verb? More specifically, are we presupposing a complete understanding on the part of the agent or are we not? If so, is it in the dialect of the agent or in the dialect of the speaker? Are we presupposing a maximal inferential capacity on the part of the agent and, if so, is it a subjective rationality or an objective rationality? It should then be obvious that a question regarding the acceptance of a particular inferential pattern in epistemic logic cannot be answered unless we answer all these other questions before.

Our theory provides a general picture in which everything appears to be a matter of degree. There is a use in which no inference is allowed. This is when we make a third-person basic atomic intentional report and do not presuppose a complete understanding of the sentence believed on the part of the agent. At the other end of the spectrum, there is a use in which the speaker makes a derived material existential report while making use of his own rationality principles. This is the quasi solipsist stance occupied by a Davidsonian radical interpreter. But in between those two extremes there is a vast array of possibilities for which our theory is able to account.

Finally, the use of a substitutional quantifier prevents the resurgence of paradoxes. The explanation goes this way. If substitutional classes may contain any expression, there appear to be strong constraints on the substitutional classes corresponding to the substitutional formulas that are meaningful. In the present case, a propositional attitude sentence with an intentional reading requires for its meaningfulness substitutional classes containing expressions with at least a linguistic meaning. When a material reading is made, the requirement is even stronger since the sentences in the substitutional class must have a content, i.e., express truth conditions or verification procedures. Now, if the propositional attitude sentence occurs in its own substitutional class (or a sentence with a quantifier ranging over the same substitutional class or presupposing the existence of such a class), it could fail to express determinate truth conditions or verification procedures. And as a matter of fact, it will not express any content whatsoever if it is a paradoxical sentence. So even if, in general, a substitutional formula can be self-referential without any danger, the ones that could become paradoxical will either fail to express a determinate content or else will satisfy constraints similar to the ones imposed by ramified type theory.

We are now in a position to see that there is sometimes something wrong in allowing that a formula be a member of its own substitutional class because, in these circumstances, we must conclude that it has truth conditions only if it is true or false, while a sentence can only be true or false if it already has truth conditions. For example, let us consider a material reading of (viii), adopt a truth conditional characterization of content, and wonder whether (viii) expresses truth conditions when it occurs as substituted in one of its substitutional instances. Since the question is precisely whether (viii) has truth conditions or not, it cannot be claimed that it is true or that it is false. And so the formula cannot tell whether it is true or false when the substituend is (viii) itself. But telling that it is true or that it is false under each of its substitutional instances was precisely the condition that had to be met in order to express truth conditions. The conclusion is that (viii) does not express truth conditions if it is allowed among the members of its own substitutional class. A moment of reflection reveals that similar problems would reoccur if the substitutional class corresponding to (viii) were to contain, perhaps not (viii) itself, but a formula with the same substitutional class as the one corresponding to (viii). The class would contain a formula that appears in its own substitutional class and we would have the same problems all over again. And if the conclusion is that it lacks truth conditions, this will again deprive (viii) itself of expressing truth conditions.

We can imagine a third case in which a substitutional class violates the meaningfulness requirement imposed above. Let us suppose that the class corresponding to (viii) contains neither (viii) itself nor a quantified formula with the same class, but a formula with a substitutional class that contains either (viii) itself or a substituend with the same class as the one corresponding to (viii). This is also a case where (i) would not become true or false under each of its substitutional instances and thus would not express definite truth conditions. As a matter of fact, we could imagine an infinite list of instances of this third case. It can be generalized as saying that a formula like (viii) does not express truth conditions if it contains a formula that involves, somewhere along the line for its definition, a substitutional class that contains a formula with the same substitutional class as the one corresponding to (viii).

The three cases examined correspond to Russell's three formulations of the vicious-circle principle, which states, first, that an entity cannot become an argument for a given function if it cannot be defined without defining the function. In this first formulation, what is ruled out is the possibility for a function to become its own argument. In its second formulation, it is said that an entity cannot be the argument for a given function if it involves a reference to the domain of the function. This rules out the possibility of having among its arguments a function that involves structurally a quantification over the same domain. In its third formulation, the principle states that an entity cannot be an argument for a given function if it presupposes that the domain of the function has already been defined and this rules out cases where the argument has (presupposes) a domain in which a function involves structurally a quantification over (or is defined by) the same domain. The conclusion is that substitutional formulas require very often for their meaningfulness substitutional classes of a determinate order. These classes must very often satisfy the constraints imposed by ramified type theory.

Of course, there are numerous differences between a substitutional language and a language similar to the one developed in *Principia Mathematica*. Our substitutional formulas do not involve subscripts indicating a specific order. They are rather systematically ambiguous expressions. Secondly, they sometimes allow for self-reference to take place. We could nonproblematically treat a formula like '(Σp) ('p' is at T and at P and is a sentence)' as meaningful and true even when the formula in question is the only one at T and at P. In other cases, ramification will be imposed only if we want to say that the sentence is meaningful (understood in the sense of expressing truth conditions or verification procedures). This is what happens in the case where we report Epimenides' belief that all Cretans are liars. The substitutional representation of such a belief report requires for its meaningfulness that the variable range over a substitutional class that does not contain the sentence itself and is well-ordered. Finally, in the case of truly paradoxical sentences, such as 'The sentence written at T and P is false', where the sentence in question is the only thing written at T and P, a substitutional representation does not block self-reference *per se,* but it enables us to show and it fails to express any content whatsoever.

MICHEL SEYMOUR

Université de Montréal

BOOK REVIEWS

Freedom within Reason. SUSAN WOLF. New York: Oxford University Press, 1990. xii, 162 p. Cloth $24.95.

Interest in perennial philosophical issues waxes and wanes for reasons it is not easy to fathom. At the moment, the topic of freedom is the focus of much attention. The rewards for those of us who are professionally committed to plow through this literature have been meager. When excessively detailed analyses and distinctions are offered as a substitute for insight and originality, the result is a clear and dreary product. There are alternatives; but they are often highly murky and elusive.

Susan Wolf's *Freedom within Reason* is a refreshing exception. Bold positions are advanced, elaborated with adequate, not exorbitant detail, and forcefully defended. The writing is crisp and not ponderous. The result is achieved in part through the virtual absence of references to current literature. Wolf's command of this literature permeates the work, but obliquely. Some would prefer more explicit engagement with the debates of the day; but the clarity and directness with which Wolf develops the line of her argument would then have been sacrificed.

Most significantly, Wolf's perspective on the issues is original. Although the theories that are surveyed are familiar, the way in which they are sorted is illuminating. Instead of the familiar classifications, Wolf offers a trichotomy: the autonomy view, the real-self view, and the reason view. The reason view, she contends, provides the best conception of the sorts of beings we must be if we are to be responsible for the results of our wills.

The asymmetric conception of responsibility she earlier introduced into the literature and here develops is also original. A responsible agent is one who has the ability to respond to the true and the good. A moral compulsive, one who is unable to do the wrong thing, is responsible if she is acting out of her perception of the rightness of her action. So in this case, determination is compatible with responsibility; but one who is (psychologically) determined to do the wrong thing lacks the ability to do otherwise and is, therefore, blameless. (For reasons I shall soon examine, *physical* determination may leave intact the ability to do otherwise.)

This conception, Wolf contends, permits a compromise stance on the debate between those who see the free-will problem as basically metaphysical and those who view it as essentially moral. She convinc-

0022-362X/92/8904/202–8

ingly rejects the antimetaphysical view identified with Peter Strawson. (Since a bona fide ability—to understand and act in accordance with the right—grounds the moral legitimacy of praise and blame, responsibility is a factual matter.) But valuational considerations enter directly into the nature of this ability: our interest in the conditions of praise and blame is an interest in knowing if someone is able to be moved to act in the right way. If, guided by the thought that morally neutral actions can be performed by a responsible agent, we ignore the valuational element and construe responsibility as some sort of robust, independent self-activity, we will not be able to explain why such agents deserve credit or merit blame.

Moral responsibility is a moral notion. And Wolf is certainly right that the link between freedom and responsibility implicates the notion of the ability to do the right thing. But that we are interested in freedom because it is relevant to the assignment of praise and blame does not imply that freedom itself must be characterized in value-laden terms. A doctor is interested in sickness because he wants to treat sick people. There is a conceptual link between sickness and treatment—only sick people should be treated. But not all sick people are treated and, more significantly, the concept of illness may well be characterizable independently of the concept of treatment. Analogously, even though only free actions deserve to be praised or blamed, not all free actions merit such responses. So freedom as such need not be defined in terms of moral concepts. If I owe Sam $10, my ability to do the right thing is just my ability to give Sam $10.

The defects of the autonomy view are familiar. Either my choice is determined, in which case I am not autonomous, or my choice has its origin in nothing, in which case it is arbitrary and in no clear sense an expression of me.

The real-self view is the compatibilist idea that, even though I am sometimes unfree when I am doing what I most want to do, there is a way of distinguishing those cases from cases of genuine freedom and responsibility in which the desires that are expressed represent the "real" self. The basic problem with this account is that the real self may be ill and, therefore, not responsible.

On what conception of responsibility can we say this? If we are claiming that the mentally ill person is not responsible for his self, we cannot substitute the autonomy conception for the real-self conception, for the autonomy conception is incoherent.

The reason view absolves a person who is ill because the person is unable to guide his action by the dictates of reason, where reason is the faculty that is inclined to lead us to the true and the good, if there are such things. Since coercion is action in which I am prevented from taking into account some relevant consideration, reason, characterized in this normative way, cannot coerce.

But is not my freedom diminished just a little if I do not autonomously choose rationality? Even if I have taken all factors into account and *A* is the rational choice, might I not be forced by reason to choose *A*?

If freedom requires that I retain the power to reject reason, I shall never have a reason to exercise this power (for if I do, the choice is rational and I am not rejecting reason). Yet even if Wolf is right, we do in fact seem to have this power and, strange as this power may be, it does not yet follow that its possession is not a necessary condition of the sort of freedom that grounds moral responsibility. Furthermore, Joel Feinberg[1] has observed that one may just delight in the realization that there are more options than one will ever want to exercise, or one may simply believe that liberty is a supreme value irreducible to its extrinsic connection to other values its possession may promote. (I shall soon examine Wolf's likely rebuttal according to which the value of liberty just is its connection to the rational or the promotion of the right.)

But if I can have a *reason* to possess a power I shall never have a reason to exercise, are we not making a crucial concession to the reason view? No, for I am not rejecting reason. I might simply believe that freedom is an independent value, even a greater value than reason, which is enhanced by the power to reject reason, even though the value of freedom depends upon a rational defense. A hedonist believes that pleasure is greater than reason, yet may concede that he must defend that position rationally. He might also take it as evident that the life of a person who derives pleasure from the knowledge that he can reject reason is enhanced by the possession of that very power.

Yet Wolf strengthens her case by a further consideration. Since the concern of the book is the discovery of the conditions of the desert of attributions of responsibility, if we concede to this current concern for autonomy, it will follow that an autonomous rational agent is more (free and) praiseworthy than a nonautonomous rational agent. But this cannot be, for there is no difference between the two agents in terms of all morally relevant considerations. Each took into account all the factors bearing on the rightness or wrongness of the act. The "freedom" of the autonomous agent is morally vacuous. True freedom (or at least the freedom that is necessary for moral responsibility) is the freedom to exercise and act in accordance with reason. A good person is deeply praiseworthy for his actions because his moral training has led him to see and to want to conform himself to the right. Similarly, a musician deserves credit for her performance because she aligns her abilities with the aesthetically relevant qualities of the performance.

---

[1] "The Interest in Liberty on the Scales," *Rights, Justice, and the Bounds of Liberty* (Princeton: University Press, 1980), pp 30–44.

Wolf is concerned to deflect criticism from those who interpret her as conflating freedom with goodness on behalf of a puritanical ethic. First, there are values other than moral ones and a rational agent should, therefore, appreciate reasons to promote nonmoral values perhaps at the expense of moral ones. Second, it is highly plausible that there are normative controversies that cannot be rationally adjudicated even by omniscient agents. If this second doctrine ("normative pluralism") is true, reason does not completely constrain values and two completely rational agents can in certain cases freely and responsibly advance conflicting goals.

Accordingly, once we abandon a narrow conception of reason, we will not be worried about the image of a person submitting grudgingly to the dictates of morality; for if his grudge has a rational basis, it will have a legitimate claim upon his attention. Reason, with its multiple sources, may recommend selfishness or adventure or the temporary suspension of morality.

Thus, Wolf supposes, the objection that true freedom is captured better by the real-self view may be put to rest. The charge that "our interest in freedom is an interest in acting according to *our own* reasons and not, as the Reason View suggests, an interest in acting by *right* reasons" (139) depends upon a failure to appreciate the variegated nature of reasons and the plausibility of normative pluralism.

Consider a test case. Alicia has received an admirable upbringing. She has internalized the values she has been taught through an open process of questioning and rational discussion. Her values track the good to the extent that this is objectively possible. Alicia finds to her dismay that she is developing powerful impulses to abandon the moral and nonmoral values she acknowledges to be correct. She finds herself identifying with these selfish impulses, while at the same time condemning the error of her potential ways. She has been taught that it is sometimes acceptable to act selfishly and to devote some of her time and energy to the promotion of her own talents and interests even when this time and energy could have been used to serve altruistic enterprises. But her selfish interests in this case are not benign; they involve gratuitous harm to others and have no redeeming quality other than Alicia's personal gratification. Alicia recognizes that her personal gratification is a reason to satisfy her selfish desires; but she also sees that it is overridden by the harm she would inflict on others. Although her struggle is a constant one, in the end she does not veer from the path of duty. She acts rightly and responsibly, but feels that life is passing her by as she presses on to do her duty at the expense of what matters most to her. She is further frustrated by the realization that she *is* doing what is overall the best thing to do—for in her special circumstances that does not coincide with the most preferred course of action.

So Alicia is moved by the right because it is right, but is neither content and perhaps may not even be able to act otherwise. She is, in a sense, a good person and may well be morally praiseworthy for her actions. But why say she is free? Alicia is unable to express her deepest self, the aspect of her being she would, if she could, display to the world as the manifestation of her individuality. (Wolf concedes the importance of these nonrational determinants of values and decisions as giving "us our individuality and distinctiveness" (146).) She is frustrated and ambivalent. Her reason tells her that *overall* she is behaving rationally; her appetitive nature, with which she identifies, could not care less.

We cannot just assume that reason will find a way out of this dilemma. To be sure, Alicia might seek psychotherapy and might discover that the elimination of her dark impulses will lead to a happier life. She might then have a reason to change. Or she might learn that the satisfaction of her seedy nature has significance sufficient to excuse at least minor harm to others. But then again she might not make these discoveries.

When Wolf concedes that nonrational determinants "at the limits . . . can be in tension with our freedom and responsibility" (146), she appears to be addressing the limitations upon our openmindedness resulting from contingencies of environment, religion, race, and the like, rather than the conflicts I am worried about in which reason is not *impaired* by one's nonrational nature, but is rather felt to be a force alien to the self. Now I am not defending the real-self view's conception of *responsibility*, for I agree that someone might be deeply insane in a way rendering him morally blameless. Yet the real self's conception of *freedom* is not thereby undermined and remains a serious competitor to Wolf's. To be sure, it is not problem-free. Wolf's reference to insanity is well-taken. A minimal level of rationality and personal integrity is required for any self to be free or responsible. In fact, if this point is conceded to Wolf—and there is no doubt that it is an important concession—the difference between the real-self and the reason views becomes very narrow. Remember that (1) the reason view does not *assert* the objectivity of reason—there may be no truth or good to discover; (2) the doctrine of normative pluralism maintains that rational adjudication of disputes is not always even theoretically possible—from a rational point of view, deep cultural impasses, for example, may have to endure; (3) reason itself imposes no limits on the sources of value or the sorts of reasons there are. In her effort to minimize the Platonic sounding character of her doctrine, Wolf has rendered the responsible agent as basically the nonirrational agent and we can all agree that irrationality and moral responsibility are incompatible.

I found Wolf's views on the connection of these ideas to determinism problematic. We are told that Rose, realizing that she should grade her ethics papers, instead succumbs to temptation and watches *The Philadelphia Story* for the seventh time. If her decision is determined by her desire to see Cary Grant again (no doubt just a *part* of the sufficient condition), she is unfree and cannot be blamed for her decision. A necessary condition of the ability to do otherwise is psychological indetermination, defined in terms of the consequent, not the antecedent factors. For example, a bleeding ulcer *physically* causes the rupture of the stomach lining, but *psychologically* causes a decision to seek medical help. (Wolf offers no account of the physical-psychological distinction, relying instead on our intuitions.) A free, psychologically undetermined action might well be physically determined. More precisely, a bodily movement which is physically necessary (in virtue of the appropriate antecedent conditions and laws) might well be the bodily movement involved in an action the agent had the ability to prevent and for which the agent may be held responsible. We cannot automatically suppose that a physical sufficient condition impedes the agent from performing an alternative action.

To a certain extent, this line of thought is part of the standard compatibilist line. Compatibilists claim that nomic necessity is different from inability, so that determination does not automatically preclude the ability to do otherwise insofar as ability is relevant to freedom and responsibility. What is distinctive in Wolf's account is her insistence that *psychological* determination is to be treated as *more* devastating than *physical* determination. Most compatibilists suppose that they will have a hard time selling their doctrine to those who believe that the most likely sort of determination in the sphere of human behavior is physical or physiological. Consequently, we are often requested to reflect on the implications of *psychological* determination because these are intuitively the most benign. How can Rose's desire render her incapacitated in relation to paper grading? Can a desire to watch Cary Grant be likened to blindness or narcolepsy? In Wolf's zeal to make psychological determinism as fundamental as far as we know as physical determinism, she seems to me to go overboard. She rejects both a strong compatibilist position according to which neither physical nor psychological determination of action per se precludes the ability to do otherwise and a modified position according to which only physical determination negates freedom. The latter view asserts that freedom is a reality if our behavior is subsumed under certain sorts of laws, if, perhaps, folk, intentionalistic, or certain forms of cognitive psychology win the day, and is an illusion if the ultimate laws of human behavior are physiological. Wolf's view, on the contrary, is that only psychological determination automatically rules out freedom and it is entirely unclear why and strongly counterintuitive that psychological determination is singled out in this way.

A clue to the reasons for Wolf's adoption of this curious opinion might be found in the examples of psychological determination provided: "Rose could have been coerced or even hypnotized into watching TV, or obsessed with Cary Grant. Alternatively, she could have a pathological aversion to grading papers that she is helpless to overcome except in the face of, say, death threats from her students or the imminent prospect of losing her job" (105). To be sure, if *all* psychological explanations are like these, Wolf is right. And given that Wolf says that "the thesis of psychological determinism has virtually nothing to support it" (101), we have some reason to believe that she is using these examples as models and drawing the warranted conclusion that not all actions are determined in these sorts of ways. But no consideration is given of the possibility of deterministic psychological theories that govern ordinary as well as compelled behavior. Since Wolf does not criticize the notion of a psychological law itself, as Donald Davidson does, I do not see why there cannot be such a law entering into a deterministic account of ordinary, noncompelled behavior.

As reviewers are wont to do, I have concentrated upon what I take to be key problems in the position. Others will, I am sure, show concern about different matters: for example, Wolf's belief that an action's being determined by a perception of its rightness is compatible with complete physical determinism is worth closer examination; her conception of rationality as the faculty most likely to lead us to the true and the good will certainly be challenged. There is enough controversy for everybody.

I would hope that Wolf will address these concerns, for it is important to know if the rationality conception of responsibility is ultimately viable. I know of no recent work that has taken us so far in the direction of establishing this result.

BERNARD BEROFSKY

Columbia University

NOTES AND NEWS

The editors regret the untimely death of Hector-Neri Castañeda, the Mahlon Powell Professor of Philosophy at Indiana University/Bloomington on September 7, 1991, at the age of 66. Born in the village of San Vicente in Guatemala, Professor Castañeda received his degrees at the Univ. of Minnesota before returning to Guatemala to teach for a year. Having studied thereafter on a fellowship at Oxford, he visited for a year at Duke before he accepted a position at Wayne State Univ. While there Professor Castañeda founded the journal *Noûs,* of which he remained the editor until the time of his death. In 1969 he moved to Indiana Univ., where he served as Dean of Latino Affairs from 1978 to 1981. Through the years he held a number of visiting appointments, directed six summer and one year-long NEH institutes, and was the recipient of a Guggenheim Fellowship. During 1979–80 he served as President of the Central Division of the APA. In 1990 he was elected to the American Academy of Arts and Sciences. Very shortly before his death he received Guatemala's highest honor, the Order del Quetzal.

In honor of Hector-Neri Castañeda, the Department of Philosophy at Indiana University has named one of its regular annual fellowships after him. In addition, a fund has been started at the Indiana Univ. Foundation. Contributions will help support Guatemalans' undergraduate or graduate study of philosophy at Indiana Univ. or at a university in Guatemala.

The editors also regret the untimely death of Gregory Vlastos, Professor Emeritus at both Princeton University and the University of California/Berkeley, on Saturday, October 12, 1991. Professor Vlastos was born in Istanbul, known then as Constantinople, coming to the United States to study at the age of 18. He earned a doctorate of divinity at Chicago Theological Seminary and a doctorate in philosophy at Harvard. Professor Vlastos taught at Queen's College in Kingston, Ontario, from 1931 to 1948, when he went to Cornell. After seven years there, he went to Princeton, where he became chairman of the Phil. Dept. Upon retiring from Princeton in 1976, he moved to Berkeley and taught there for more than a decade. During his career, Professor Vlastos wrote several texts and edited more than 50 books. He served as President of the Eastern Division of the APA. Among his numerous awards and honors, he won two Guggenheim Fellowships and was awarded the MacArthur Foundation fellowship in 1990.

In honor of Gregory Vlastos, a conference on the theme "Socratic Studies" will be held at the University of California/Berkeley on May 1–3. During the weekend there will also be a memorial service. The conference will include five sessions: (1) Socratic Method; (2) Virtue; (3) Political Theory; (4) Love and Sexuality; (5) The Theory of Forms. The speakers will be Julia Annas, Alan Code, Gail Fine, David Halperin, Terence Irwin, Richard Kraut, Jonathan Lear, Robert Meister, Alexander Nehamas, Martha Nussbaum, Gerasimos Santas, and Bernard Williams. Those interested in attending the memorial service and conference may contact Robert Meister, Politics Dept., Kresge College, Univ. of California, Santa Cruz, CA 95064.

The Department of Philosophy at the University of California/Riverside is pleased to announce the appointment of Jonathan Wilwerding as Assistant Professor, Mark Ravizza as Assistant Professor, Genoveva Marti as Assistant Professor, and H Carl Hoefer as Assistant Professor.

The Department of Philosophy at Emory University is pleased to announce the appointment of David Carr as Professor.

The Department of Philosophy at Ohio State University is pleased to announce the appointments of Neil Tennant and Mark Wilson as Professors, William Taschek and Kathleen Cook as Assistant Professors, and Anna Greco as Instructor.

The Department of Philosophy at Purdue University is pleased to announce the appointment of Leonard Harris as Professor of Philosophy and Director of the Afro-American Studies and Research Center, and Martin J. Matustik as Assistant Professor of Philosophy.

NEW BOOKS

ACKERMAN, BRUCE: *We The People, Vol. I: Foundations*. Cambridge. Harvard, 1991. x, 369 p. Cloth $24.95.

ALSTON, WILLIAM P.: *Perceiving God: The Epistemology of Religious Experience*. Ithaca: Cornell, 1991. xii, 320 p. Cloth $36.95.

ANGLIN, W. S.: *Free Will and the Christian Faith*. New York: Oxford, 1991, viii, 218 p. Cloth $55.00.

AUNE, BRUCE: *Knowledge of the External World*. New York· Routledge, 1991. xiv, 242 p. Cloth $39.95.

AYERS, MICHAEL: *Locke: Vol. I, Epistemology; Vol. II, Ontology*. New York: Routledge, 1991. x, 341 p.; vi, 341 p. Cloth $127.50, 2 vols.

BAYNES, KENNETH: *The Normative Grounds of Social Criticism: Kant, Rawls, and Habermas*. Albany: SUNY, 1991. xi, 242 p. Paper $18.95.

BENJAMIN, ANDREW: *Art, Mimesis, and the Avant-Garde: Aspects of a Philosophy of Difference*. New York: Routledge, 1991. ix, 217 p. Cloth $49.95, paper $15.95.

BEYLEVELD, DERYCK: *The Dialectical Necessity of Morality*. Chicago: University Press, 1992. xxxviii, 523 p. Cloth $55.00, paper $22.50.

BIGELOW, JOHN, and ROBERT PARGETTER: *Science and Necessity*. New York: Cambridge, 1991. ix, 410 p. Cloth $49.50.

BLUSTEIN, JEFFREY: *Care and Commitment: Taking the Personal Point of View*. New York: Oxford, 1991. x, 273 p. Cloth $35 00.

BOOTHBY, RICHARD: *Death and Desire: Psychoanalytic Theory in Lacan's Return to Freud*. New York: Routledge, 1991. xv, 266 p. Cloth $42.50, paper $15.95.

BRAATEN, JANE: *Habermas' Critical Theory of Society*. Albany: SUNY, 1991. x, 191 p. Cloth $44.50, paper $14.95.

BROOKE, JOHN HEDLEY: *Science and Religion: Some Historical Perspectives*. New York: Cambridge, 1991. x, 422 p. Cloth $44.50, paper $12 95.

BROWN, JAMES ROBERT: *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. New York: Routledge, 1991. xi, 175 p. Cloth $49.95.

BUCHANAN, ALLEN: *Secession: The Morality of Political Divorce from Fort Sumter to Lithuania and Quebec*. Boulder, CO: Westview, 1991. xviii, 174 p. Cloth $38.50, paper $14.95.

CLARK, STEPHEN R. L.: *God's World and the Great Awakening: Limits and Renewals 3*. New York: Oxford, 1991. viii, 246 p. Cloth $55.00.

CLARK, S. H.: *Paul Ricoeur*. New York: Routledge, 1991. viii, 216 p. Cloth $52.50, paper $19.95.

CORNELL, DRUCILLA: *Beyond Accommodation: Ethical Feminism, Deconstruction, and the Law*. New York: Routledge, 1991. ix, 239 p. Cloth $56.50, paper $14.95.

COWLEY, FRASER: *Metaphysical Delusion*. Buffalo: Prometheus, 1991. 200 p. Cloth $29.95.

CUMMING, ROBERT DENOON: *Phenomenology and Deconstruction, Vol. 1: The Dream is Over*. Chicago University Press, 1991. 255 p. Cloth $49.95, paper $17.95.

DARDEN, LINDLEY: *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford, 1991. xi, 314 p. Cloth $39.95.

DENNETT, DANIEL C : *Consciousness Explained*. New York: Little & Brown, 1991. xiii, 511 p. Cloth $27.95.

DIAMOND, CORA: *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. Cambridge: MIT, 1991. xi, 396 p. Cloth $32.50.

DUMMETT, MICHAEL· *Frege: Philosophy of Mathematics*. Cambridge: Harvard, 1991. xiii, 331 p. Cloth $34 95.

EELS, ELLERY: *Probabilistic Causality*. New York: Cambridge, 1991. xii, 413 p. Cloth $44.50

EMMANUEL, EZEKIEL J.: *The Ends of Human Life: Medical Ethics in a Liberal Polity*. Cambridge: Harvard, 1992 x, 307 p. Cloth $34.95.

EVNINE, SIMON· *Donald Davidson*. Stanford: University Press, 1991. viii, 198 p. Cloth $32.50, paper $10 95.

EZORSKY, GERTRUDE: *Racism and Justice: The Case for Affirmative Action*. Ithaca: Cornell, 1991. ix, 140 p. Cloth $21.95, paper $6.95.

FARNESS, JAY: *Missing Socrates: Problems of Plato's Writing*. University Park: Penn State UP, 1991. x, 206 p. Cloth $29.50.

FEENBERG, ANDREW: *Critical Theory of Technology*. New York: Oxford, 1991. xi, 235 p. Cloth $35.00, paper $14.95.

FERREIRA, M. JAMIE: *Transforming Vision: Imagination and Will in Kierkegaardian Faith*. New York: Oxford, 1991. 168 p. Cloth $39.95.

FISCH, MENACHEM: *William Whewell, Philosopher of Science*. New York: Oxford, 1991. xiv, 220 p. Cloth $59.00.

FISCHMANN, DENNIS: *Political Discourse in Exile: Karl Marx and the Jewish Question*. Amherst: Massachusetts UP, 1991. viii, 160 p Cloth $21.95.

FOSTER, JOHN: *The Immaterial Self: A Defense of the Cartesian Dualist Conception of the Mind*. New York: Routledge, 1991. ix, 298 p. Cloth $49.95.

FUKUYAMA, FRANCIS: *The End of History and the Last Man*. New York: Free Press, 1992. xxiii, 418 p. Cloth $24.95.

GAITA, RAIMOND: *Good & Evil: An Absolute Conception*. New York: St. Martin's, 1991. ix, 340 p. Cloth $49.95.

GALE, RICHARD M.: *On the Nature and Existence of God*. New York: Cambridge, 1991 viii, 422 p. Cloth $44.50.

GALSTON, WILLIAM A.: *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State*. New York: Cambridge, 1991. xi, 343 p. Cloth $49.50, paper $16.95.

GANE, MIKE: *Baudrillard: Critical and Fatal Theory*. New York: Routledge, 1991. 243 p. Cloth $49.95, paper $15.95.

GATENS, MOIRA: *Feminism and Philosophy: Perspectives on Difference and Equality*. Bloomington: Indiana UP, 1991. 162 p. Cloth $39.95, paper $14.95.

GAY, PETER: *Reading Freud: Explorations and Entertainments*. New Haven: Yale, 1991. xvi, 204 p. Cloth $28.00, paper $15.00.

GELVEN, MICHAEL: *Why Me? A Philosophical Inquiry into Fate*. DeKalb: Northern Illinois UP, 1991. ix, 199 p. Cloth $22.00.

GOLDMAN, ALVIN I.: *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge: MIT, 1992. viii, 336 p. Cloth $39.95.

GOODKIN, RICHARD E.: *The Tragic Middle: Racine, Aristotle, Euripides*. Madison: Wisconsin UP, 1991. ix, 211 p. Cloth $37.00.

GRICE, PAUL: *The Conception of Value*. New York: Oxford, 1991. 164 p. Cloth $45.00.

GRIM, PATRICK: *The Incomplete Universe: Totality, Knowledge, and Truth*. Cambridge: MIT, 1991. viii, 165 p. Cloth $25.00

HARRISON, BERNARD: *Inconvenient Fictions: Literature and the Limits of Theory*. New Haven: Yale, 1991. viii, 293 p. Cloth $35.00.

HAWTHORN, GEOFFREY: *Plausible Words: Possibility and Understanding in History and the Social Sciences*. New York: Cambridge, 1991. xiii, 192 p. Cloth $44.50.

HILL, CLAIRE ORTIZ: *Word and Object in Husserl, Frege, and Russell: The Roots of Twentieth-century Philosophy*. Athens: Ohio UP, 1991. xv, 213 p. Cloth $34.95.

HILL, THOMAS E. JR.: *Autonomy and Self-Respect*. New York. Cambridge, 1991. ix, 218 p. Cloth $39.50, paper $12.95.

HODGSON, DAVID: *The Mind Matters: Consciousness and Choice in a Quantum World*. New York: Oxford, 1991 ix, 484 p. Cloth $89.00.

HOLUB, ROBERT C.: *Jürgen Habermas: Critic in the Public Sphere*. New York: Routledge, 1991. xii, 210 p. Cloth $49.95, paper $14.95.

HOULGATE, STEPHEN: *Freedom, Truth, and History: An Introduction to Hegel's Philosophy*. New York: Routledge, 1991. xviii, 263 p. Cloth $49.95, paper $16.95.

JARDINE, NICHOLAS: *The Sciences of Inquiry: On the Reality of Questions in the Sciences*. New York: Oxford, 1991. x, 245 p. Cloth $55.00.

JOHNSTONE, ALBERT A.: *Rationalized Epistemology: Taking Solipsism Seriously*. Albany: SUNY, 1991. xvii, 361 p. Cloth $59.50, paper $19.95.

KAUFMAN-OSBORN, TIMOTHY V : *Politics/Sense/Experience: A Pragmatic Inquiry into the Promise of Democracy*. Ithaca: Cornell, 1991. xviii, 327 p. Cloth $21.95.

KOHNKE, KLAUS CHRISTIAN: *The Rise of Neo-Kantianism: German Academic Philosophy between Idealism and Positivism*. New York: Cambridge, 1992. xiii, 290 p. Cloth $64.95.

KLEINIG, JOHN: *Valuing Life: Studies in Moral, Political, and Legal Philosophy*. Princeton: University Press, 1991. xvii, 284 p. Cloth $35.00.

KUHNS, RICHARD: *Tragedy: Contradiction as Repression*. Chicago: University Press, 1991. x, 180 p. Cloth $27.50.

KUPPERMAN, JOEL J.: *Character*. New York: Oxford, 1991. vi, 193 p. Cloth $28.00.

LANG, BEREL: *Writing and the Moral Self*. New York: Routledge, 1991. ix, 179 p. Cloth $42.50, paper $13.95.

LEIBER, JUSTIN: *An Invitation to Cognitive Science*. Cambridge: Blackwell, 1991. x; 167 p. Cloth $49.95, paper $17.95.

LEVI, ISAAC: *The Fixation of Belief and Its Undoing*. New York: Cambridge, 1992. x, 192 p. Cloth $34.50.

LIEB, IRWIN: *Past, Present, and Future: A Philosophic Essay about Time*. Champaign: Illinois UP, 1992. viii, 260 p. Cloth $35.00, paper $14.95.

LIPMAN, MATTHEW: *Thinking in Education*. New York: Cambridge, 1991. ix, 280 p. Cloth $49.50, paper $19.95.

LLOYD, G. E. R.: *Methods and Problems in Greek Science*. New York: Cambridge, 1991. xiv, 457 p. Cloth $59.50.

LOUX, MICHAEL J.: *Primary Ousia: An Essay on Aristotle's Metaphysics Z and H*. Ithaca: Cornell, 1991. xii, 285 p. Cloth $39.95.

LUKES, STEVEN: *Moral Conflict and Politics*. New York: Oxford, 1991. ix, 323 p. Cloth $69.00.

MCBRIDE, WILLIAM L.: *Sartre's Political Theory*. Bloomington: Indiana UP, 1991. 246 p. Cloth $37.50, paper $12.95.

MALCOLM, JOHN: *Plato on the Self-predication of Forms*. New York: Oxford, 1991. 231 p. Cloth $55.00.

MANN, PAUL: *The Theory-death of the Avant-Garde*. Bloomington: Indiana UP, 1991. 154 p. Cloth $29.95.

MELLOR, D. H.: *Matters of Metaphysics*. New York: Cambridge, 1991. xx, 295 p. Cloth $44.50.

MILLER, RICHARD B.: *Interpretations of Conflict: Ethics, Pacifism, and the Just-war Tradition*. Chicago: University Press, 1991. x, 294 p. Cloth $48.00, paper $17.95.

MULHOLLAND, LESLIE A.: *Kant's System of Rights*. New York: Columbia, 1990. xvii, 434 p. Cloth $49.00.

NEVILLE, ROBERT CUMMINGS: *A Theology Primer*. Albany: SUNY, 1991. xxix, 221 p. Cloth $39.50, paper $12.95.

NORTON, ROBERT E.: *Herder's Aesthetics and the European Enlightenment*. Ithaca: Cornell, 1991. xi, 257 p. Cloth $35.95.

OLSSON, GUNNAR: *Lines of Power: Limits of Language*. Minneapolis. Minnesota UP, 1991. 229 p. Cloth $19.95

OPHIR, ADI: *Plato's Invisible Cities: Discourse and Power in the* Republic. Savage, MD: Barnes & Noble, 1991. 211 p. Cloth $47.50.

PARRY, WILLIAM T., and EDWARD A. HACKER: *Aristotelian Logic*. Albany: SUNY, 1991. ix, 545 p. Cloth $49.50, paper $16.95.

PEARSON, KEITH-ANSELL: *Nietzsche Contra Rousseau: A Study of Nietzsche's Moral and Political Thought*. New York: Cambridge, 1992. xvii, 284 p. Cloth $49.95.

PERRIN, RON: *Max Scheler's Concept of the Person: An Ethics of Humanism*. New York: St Martin's, 1991. xi, 156 p. Cloth $35.00.

PLETSCH, CARL: *Young Nietzsche: Becoming a Genius*. New York: Free Press, 1991. x, 261 p. Cloth $22.95.

POOLE, ROSS: *Morality and Modernity*. New York: Routledge, 1991. xii, 196 p. Cloth $49.95, paper $16.95.

PORTER, ELISABETH J.: *Women and Moral Identity*. Concord, MA: Allen & Unwin, 1991. xiii, 223 p. Paper $22.95.

RAJCHMAN, JOHN: *Truth and Eros: Foucault, Lacan, and the Question of Ethics*. New York: Routledge, 1991. vi, 154 p. Cloth $42.50, paper $13.95.

RAKOWSKI, ERIC: *Equal Justice*. New York: Oxford, 1991. xii, 385 p. Cloth $49.95.

RAY, CHRISTOPHER: *Time, Space, and Philosophy*. New York: Routledge, 1991. xi, 268 p. Cloth $55.00, paper $16.95.

RESCHER, NICHOLAS: *A System of Pragmatic Idealism, Vol. I: Human Knowledge in Idealistic Perspective*. Princeton: University Press, 1992. xiv, 335 p. Cloth $37.50.

ROCKEFELLER, STEVEN C.: *John Dewey: Religious Faith and Democratic Humanism*. New York: Columbia, 1991. xiii, 683 p. Cloth $50.00.

ROSENTHAL, SANDRA B., and PATRICK L. BOURGEOIS: *Mead and Merleau-Ponty: Toward a Common Vision*. Albany: SUNY, 1991. ix, 231 p Cloth $49.50, paper $16.95.

ROWE, WILLIAM L.: *Thomas Reid on Freedom and Morality*. Ithaca: Cornell, 1991. x, 189 p. Cloth $26.95.

SAINSBURY, MARK: *Logical Forms: A Introduction to Philosophical Logic*. Cambridge: Blackwell, 1991. vii, 398 p. Cloth $64.95, paper $21.95.

SCHIAPPA, EDWARD: *Protagoras and Logos: A Study in Greek Philosophy and Rhetoric*. Columbia: South Carolina UP, 1991. xvii, 239 p. Cloth $29.95.

SAWICKI, JANA: *Disciplining Foucault: Feminism, Power, and the Body*. New York: Routledge, 1991. xi, 130 p. Cloth $42.50, paper $13.95.

SCHICK, FREDERIC: *Understanding Action: An Essay on Reasons*. New York: Cambridge, 1991. viii, 167 p. Cloth $39.50, paper $11.95.

SCHOPP, ROBERT F.: *Automatism, Insanity, and the Psychology of Criminal Responsibility: A Philosophical Inquiry*. New York: Cambridge, 1991. xi, 277 p. Cloth $39.50.

SEBEOK, THOMAS A.: *Semiotics in the United States*. Bloomington: Indiana UP, 1991. 173 p. Cloth $29.95, paper $10 95.

SEGAL, JEROME M: *Agency and Alienation: A Theory of Human Presence*. Savage, MD: Rowman & Littlefield, 1991. xii, 259 p. Cloth $42.50.

SENCHUK, DENNIS M.: *Against Instinct: From Biology to Philosophical Psychology*. Philadelphia: Temple, 1991. xxii, 277 p. Cloth $44 95

SHANKS, ANDREW: *Hegel's Political Theology*. New York: Cambridge, 1992. xiii, 234 p. Cloth $49.50.

SHAPIRO, GARY: *Alcyone: Nietzsche on Gifts, Noise, and Women*. Albany: SUNY, 1991. xii, 158 p. Cloth $39 50, paper $12.95.

SHER, GILA: *The Bounds of Logic: A Generalized Viewpoint*. Cambridge: MIT, 1991. xv, 178 p. Cloth $27.50.

TYE, MICHAEL: *The Imagery Debate*. Cambridge: MIT, 1991. xiv, 172 p. Cloth $22.50.

VAN FRAASSEN, BAS C.: *Quantum Mechanics: An Empiricist View*. New York: Oxford, 1991. xvi, 359 p. Cloth $69.00, paper $29 95.

VARELA, FRANCISCO J., EVAN THOMSON, and ELEANOR ROSCH: *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT, 1991. xx, 308 p. Cloth $25.00.

VERENE, DONALD PHILLIP: *The New Art of Autobiography: An Essay on the Life of Giambattista Vico Written by Himself*. New York: Oxford, 1991. xi, 263 p. Cloth $59.00.

VESEY, GODFREY: *Inner and Outer: Essays on a Philosophical Myth*. New York: St. Martin's, 1991. viii, 258 p. Cloth $45.00.

WAGMAN, MORTON: *Cognitive Science and Concepts of Mind: Toward a General Theory of Human and Artificial Intelligence*. New York: Praeger, 1991. xvi, 192 p. Cloth $45.00.

WAJCMAN, JUDY: *Feminism Confronts Technology*. University Park: Penn State UP, 1991. x, 184 p. Cloth $28.95, paper $13.95.

WALLWORK, ERNEST: *Psychoanalysis and Ethics*. New Haven: Yale, 1991. xiii, 344 p. Cloth $35.00.

WALTON, DOUGLAS N.: *Begging the Question: Circular Reasoning as a Tactic of Argumentation*. Westport, CT: Greenwood, 1991. xv, 340 p. Cloth $49.95.

WERHANE, PATRICIA H.: *Adam Smith and His Legacy for Modern Capitalism*. New York: Oxford, 1991. ix, 219 p. Cloth $29 95

WHITE, STEPHEN L.: *The Unity of Self*. Cambridge: MIT, 1991. xii, 424 p. Cloth $35.00.

WINFIELD, RICHARD DIEN: *Freedom and Modernity*. Albany: SUNY, 1991 xvi, 313 p. Cloth $44.50, paper $14.95.

WREN, THOMAS E.: *Caring about Morality: Philosophical Perspectives in Moral Psychology*. Cambridge: MIT, 1991. x, 202 p. Cloth $25.00.

YOURGRAU, PALLE: *The Disappearance of Time: Kurt Gödel and the Idealistic Tradition in Philosophy*. New York: Cambridge, 1992. x, 182 p. Cloth $39.50.

YERUSHALMI, YOSEF HAYIM: *Freud's Moses: Judaism Terminable and Interminable*. New Haven: Yale, 1991. xix, 159 p. Cloth $25.00.

ZUIDERVAART, LAMBERT: *Adorno's Aesthetic Theory: The Redemption of Illusion*. Cambridge: MIT, 1991. xxiv, 388 p. Cloth $37.50.

NEW BOOKS: ANTHOLOGIES

AMSELEK, PAUL, and NEIL MACCORMICK, eds.: *Controversies about Law's Ontology*. Edinburgh: University Press, 1991. xiii, 160 p. Cloth $45.00.

ANSELL-PEARSON, KEITH, ed.: *Nietzsche and Modern German Thought*. New York: Routledge, 1991. vi, 314 p. Cloth $49.95.

BAIRD, ROBERT M., and STUART E. ROSENBAUM, eds.: *Pornography: Private Right or Public Menace?* Buffalo: Prometheus, 1991. 248 p. Paper $14.95.

BARKER, PETER, and ROGER ARIEW, eds.: *Studies in Philosophy and the History of Philosophy, Vol. 24: Revolution and Continuity*. Washington: Catholic UP, 1991. v, 222 p. Cloth $42.95.

BARTLEY, W. W. III, and STEPHEN KRESGE, eds.: *The Collected Works of F. A. Hayek, Vol. 3: The Trend of Economic Thinking*. Chicago: University Press, 1991. xi, 388 p. Cloth $40.00.

BENDER, JOHN, and DAVID E. WELLBERY, eds.: *Chronotypes: The Construction of Time*. Stanford: University Press, 1991. xi, 258 p. Cloth $37.50, paper $12.95.

BENEDIKT, MICHAEL L., ed.: *Cyberspace: First Steps*. Cambridge: MIT, 1991. vii, 304 p. Cloth $24.95.

BERTOTTI, B., et al., eds.: *Modern Cosmology in Retrospect*. New York: Cambridge, 1991 xx, 426 p. Cloth $69.50.

BLUMENTHAL, HENRY, and HOWARD ROBINSON, eds.: *Aristotle and the Later Tradition*. New York: Oxford, 1991. x, 277 p. Cloth $65 00

BORMANN, F. HERBERT, and STEPHEN R. KELLERT, eds.: *Ecology, Economics, Ethics: The Broken Circle*. New Haven: Yale, 1991. xviii, 233 p. Cloth $26.50.

BOYD, RICHARD, PHILIP GASPER, and J. D. TROUT, eds.: *The Philosophy of Science*. Cambridge: MIT, 1991. xvi, 816 p. Cloth $50.00, paper $25.00.

BRADY, JAMES B., and NEWTON GARVER, eds.: *Justice, Law, and Violence*. Philadelphia: Temple, 1991. x, 281 p. Cloth $44.95.

BURCHELL, GRAHAM, COLIN GORDON, and PETER MILLER, eds.: *The Foucault Effect: Studies in Governmentality*. Chicago: University Press, 1991. x, 307 p. Cloth $58.00, paper $19.95

BURNS, J. H., ed.: *The Cambridge History of Political Thought 1450–1700*. New York: Cambridge, 1991. xii, 798 p. Cloth $120.00.

BUTTON, GRAHAM, ed : *Ethnomethodology and the Human Sciences*. New York: Cambridge, 1991. xiii, 278 p. Cloth $54.50, paper $19.95.

CALHOUN, CRAIG, ed.: *Habermas and the Public Sphere*. Cambridge: MIT, 1992. x, 498 p. Cloth $45.00.

CARVER, TERRELL, ed.· *The Cambridge Companion to Marx*. New York. Cambridge, 1992. xiii, 357 p. Cloth $59 95, paper $17.95.

CAWS, MARY ANN, ed.: *City Images: Perspectives from Literature, Philosophy, and Film*. New York: Gordon & Breach, 1991 278 p. Paper $15.00.

CORLETT, J. ANGELO, ed.: *Equality and Liberty: Analyzing Rawls and Nozick*. New York: St. Martin's, 1991. x, 397 p. Cloth $39.95.

CORNELL, DRUCILLA, et al., eds.: *Hegel and Legal Theory*. New York: Routledge, 1991. xviii, 359 p. Cloth $45.00, paper $15.95.

CUMMINS, ROBERT, and JOHN POLLOCK, eds.: *Philosophy and AI: Essays at the Interface*. Cambridge: MIT, 1992. xi, 304 p. Cloth $29.95.

DASCAL, MARCELO, ed.: *Cultural Relativism and Philosophy: North and Latin American Perspectives*. New York: Brill, 1991. xii, 316 p. Cloth $74.36.

EELS, ELLERY, and TOMASZ MARUSZEWSKI, eds.: *Probability and Rationality: Studies on Jonathan Cohen's Philosophy of Science*. Atlanta: Rodopi, 1991. 345 p. Cloth $75.00.

ELSTER, JOHN, and JOHN E. ROEMER, eds.: *Interpersonal Comparisons of Well-being*. New York: Cambridge, 1991. viii, 400 p. Cloth $49.50.

ENGEL, J. RONALD, and JOAN GIBB ENGEL, eds.: *Ethics of Environment and Development: Global Challenge, International Response*. Tucson: Arizona UP, 1991. xv, 264 p. Cloth $29.95, paper $14.95.

EVERSON, STEPHEN, ed.: *Companions to Ancient Thought 2: Psychology*. New York: Cambridge, 1991. vii, 269 p. Cloth $54.50.

FREADMAN, RICHARD, and LLOYD REINHARDT, eds.: *On Literary Theory and Philosophy: A Cross-disciplinary Encounter*. New York: St. Martin's, 1991. xi, 281 p. Cloth $45.00.

FREY, R. G., and CHRISTOPHER W. MORRIS, eds.: *Violence, Terrorism, and Justice*. New York: Cambridge, 1991. x, 319 p. Cloth $54.50, paper $17.95.

FREY, R. G., and CHRISTOPHER W. MORRIS, eds.: *Liability and Responsibility: Essays in Law and Morals*. New York: Cambridge, 1991. x, 430 p. Cloth $54.50.

GREENWOOD, JOHN D., ed.: *The Future of Folk Psychology: Intentionality and Cognitive Science.* New York: Cambridge, 1992. ix, 293 p. Cloth $44.50, paper $12.95.

HILEY, DAVID, JAMES BOHMAN, and RICHARD SCHUSTERMAN, eds : *The Interpretive Turn: Philosophy, Science, Culture.* Ithaca: Cornell, 1992 x, 322 p Cloth $39.95, paper $12.95.

HORTON, JOHN, and SUSAN MENDUS, eds.: *John Locke: A Letter Concerning Toleration in Focus.* New York: Routledge, 1991. vii, 196 p. Cloth $59.95, paper $16.95.

HUYSSEN, ANDREAS, and DAVID BATHRICK, eds.: *Modernity and the Text: Revisions of German Modernism.* New York: Columbia, 1991. vii, 244 p. $17.50.

JACKSON, FRANK, ed.: *Conditionals.* New York: Oxford, 1991. 238 p. Cloth $39.95, paper $14.95.

JECKER, NANCY S., ed.: *Aging and Ethics: Philosophical Problems in Gerontology.* Clifton, NJ: Humana, 1991. xi, 394 p. Cloth $32.50.

KELLEY, DONALD R., ed.: *Versions of History: From Antiquity to the Enlightenment.* New Haven: Yale, 1991. xii, 515 p. Cloth $45.00, paper $18 00.

KELLY, PATRICK HYDE,, ed.: *Locke on Money, Vol. 1.* New York: Oxford, 1991. xv, 342 p. Cloth $89.00.

KOSUTH, JOSEPH: *Art after Philosophy and After: Collected Writings, 1966–1990.* GABRIELE GUERCIO, ed. Cambridge: MIT, 1991. xlii, 289 p. Cloth $30.00.

LASH, SCOTT, and JONATHAN FRIEDMAN, eds.: *Modernity and Identity.* Cambridge: Blackwell, 1991. viii, 379 p. Cloth $64.95, paper $19.95.

LEHRER, KEITH, and ERNEST SOSA, eds.: *The Opened Curtain: A U.S. Soviet Philosophy Summit.* Boulder: Westview, 1991. ix, 308 p. Paper $38.50.

LORD, CARNES, and DAVID K. O'CONNOR, eds.. *Essays on the Foundation of Aristotelian Political Science.* Berkeley: California UP, 1991. 255 p. Cloth $45.00.

MCLAUGHLIN, BRIAN, ed.: *Dretske and His Critics.* Cambridge: Blackwell, 1991. xvii, 231 p. Cloth $47.95.

MAY, LARRY, and STACEY HOFFMAN, eds.: *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics.* Savage, MD: Rowman & Littlefield, 1991. viii, 292 p. Cloth $50.00, paper $17.95.

MEEKS, GAY, ed.: *Thoughtful Economic Man: Essays on Rationality, Moral Rules, and Benevolence.* New York: Cambridge, 1991. ix, 160 p. Cloth $29.95

MILLER, DAVID, ed.: *Liberty.* New York: Oxford, 1991. 223 p. Cloth $58.00, paper $18.95.

NEU, JEROME, ed.· *The Cambridge Companion to Freud.* New York: Cambridge, 1992. viii, 356 p. Cloth $59.95, paper $17.95.

OGILVY, JAMES, ed.: *Revisioning Philosophy.* Albany: SUNY, 1992. xxiii, 318 p. Paper $10.95.

OLDENQUIST, ANDREW, and MENACHEM ROSNER, eds.: *Alienation, Community, and Work*. Westport: Greenwood, 1991. viii, 211 p. Cloth $42.95.

OSLER, MARGARET J., ed.: *Atoms, Pneuma, and Tranquility: Epicurean and Stoic Themes in European Thought*. New York: Cambridge, 1991. xii, 304 p. Cloth $49.50.

PARKES, GRAHAM, ed.: *Nietzsche and Asian Thought*. Chicago: University Press, 1991. xii, 253 p. Cloth $27.50

PAUL, ELLEN FRANKEL, FRED D. MILLER, and JEFFREY PAUL, eds.: *Ethics, Politics, and Human Nature*. Cambridge Blackwell, 1991 ix, 191 p. Paper $16.95.

REAMER, FREDERIC G., ed.: *AIDS and Ethics*. New York: Columbia, 1991. xviii, 317 p. Cloth $29.50.

ROSENBLUM, NANCY L., ed.: *Liberalism and the Moral Life*. Cambridge: Harvard, 1991. 302 p. $14.95.

ROSENTHAL, DAVID M., ed.: *The Nature of Mind*. New York: Oxford, 1991. x, 642 p. Cloth $49.95, paper $19.95.

ROSSI, PHILIP J., and MICHAEL WREN, eds.: *Kant's Philosophy of Religion Reconsidered*. Bloomington: Indiana UP, 1991. xxiii, 205 p. Cloth $25.00.

ROUNER, LEROY, ed.: *On Community*. Notre Dame: University Press, 1991. xvi, 212 p. Cloth $29.95.

RUSSELL, STUART, and ERIC WEFALD, eds.: *Do the Right Thing: Studies in Limited Rationality*. Cambridge: MIT, 1991. xx, 200 p. Cloth $27.50.

SAUNDERS, SIMON, and HARVEY R. BROWN, eds.: *The Philosophy of Vacuum*. New York: Oxford, 1991. viii, 291 p. Cloth $65.00.

SCHOEDINGER, ANDREW B., ed.: *The Problem of Universals*. Atlantic Highlands, NJ: Humanities, 1991. x, 360 p. Cloth $60.00.

SILVERMAN, HUGH J., ed.: *Gadamer and Hermeneutics: Science, Culture, Literature*. New York: Routledge, 1991. 332 p. Cloth $45.00, paper $15.95.

SOLOMON, ROBERT C., and KATHLEEN M. HIGGINS, eds.: *The Philosophy of (Erotic) Love*. Lawrence: Kansas UP, 1991. xii, 522 p. Cloth $45.00, paper $17.95.

VALLENTYNE, PETER, ed.: *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*. New York. Cambridge, 1991. xvi, 339 p. Cloth $59.50, paper $18.95.

WOLIN, RICHARD, ed.: *The Heidegger Controversy: A Critical Reader*. New York: Columbia, 1991. x, 315 p. Cloth $35.00.

NEW BOOKS: PAPERBACKS

BELL, DAVID: *Husserl*. New York: Routledge, 1992. xv, 266 p. $16.95.

DEVITT, MICHAEL: *Realism and Truth*, 2nd Ed. Cambridge: Blackwell, 1991. xii, 327 p. $27.95.

GELLNER, ERNEST: *Spectacles and Predicaments: Essays in Social Theory*. New York: Cambridge, 1991. viii, 385 p. $19.95.

GRICE, PAUL: *Studies in the Way of Words*. Cambridge Harvard, 1991. viii, 394 p. $14.95.

HAMPSHIRE, STUART: *Innocence and Experience*. Cambridge: Harvard, 1992. 195 p. $9.95.

KAGAN, SHELLY. *The Limits of Morality*. New York: Oxford, 1991. xii, 415 p. $24.95.

KERNAN, ALVIN: *The Death of Literature*. New Haven: Yale, 1992. ix, 230 p. $10.00.

KIVY, PETER: *Music Alone: Philosophical Reflections on the Purely Musical Experience*. Ithaca: Cornell, 1991. xii, 226 p. $10.95.

KRAUT, RICHARD: *Aristotle on the Human Good*. Princeton: University Press, 1991. viii, 379 p. $16.95.

MACEDO, STEPHEN: *Liberal Virtues: Citizenship, Virtue, and Community in Liberal Constitutionalism*. New York: Oxford, 1991. viii, 306 p. $15.95.

MARTIN, ROBERT M.: *The Philosopher's Dictionary*. Lewiston, NY: Broadview, 1991. 180 p. $10.95.

MEYERS, DIANA T.: *Self, Society, and Personal Choice*. New York: Columbia, 1991. xiii, 287 p. $19.50.

POSTER, MARK: *The Mode of Information: Poststructuralism and Social Context*. Chicago: University Press, 1992. vii, 179 p. $16.95.

SCHOEDINGER, ANDREW B., ed.: *Introduction to Metaphysics: The Fundamental Questions*. Buffalo: Prometheus, 1991. 469 p. $21 95.

SHEETS-JOHNSTONE, MAXINE: *The Roots of Thinking*. Philadelphia: Temple, 1992. ix, 389 p. $19.95.

SUTER, RONALD: *Interpreting Wittgenstein: A Cloud of Philosophy, a Drop of Grammar*. Philadelphia: Temple, 1992. xvi, 256 p. $22.95.

SWINBURNE, RICHARD: *The Existence of God*, Rev. Ed. New York: Oxford, 1991 327 p. $26 00.

VERENE, DONALD PHILLIP: *Vico's Science of Imagination*. Ithaca: Cornell, 1992. 227 p. $10.95.

WARREN, MARK: *Nietzsche and Political Thought*. Cambridge: MIT, 1991. xv, 311 p. $12.50.

NEW BOOKS: TRANSLATIONS

ALBERTI, LEON BATTISTA: *On the Art of Building in Ten Books*. JOSEPH RYKWERT, NEIL LEACH, and ROBERT TAVERNOR, trans. Cambridge: MIT, 1991. 472 p. Cloth $47.50, paper $16.95.

ALEXANDER OF APHROSISIAS: *On Aristotle's Prior Analytics 1.1–7*. JONATHAN BARNES et al. trans. Ithaca: Cornell, 1991 vii, 252 p. Cloth $47.95.

ALLEN, R. E., trans.: *The Dialogues of Plato: Volume II: The Symposium*. New Haven: Yale, 1991. xi, 178 p. Cloth $25.00.

AMMONIUS: *On Aristotle's Categories*. S. MARC COHEN and GARETH B. MATTHEWS, trans. Ithaca: Cornell, 1991. 170 p. Cloth $46.50.

ARMSTRONG, T. J., ed. and trans.: *Michel Foucault, Philosopher*. New York: Routledge, 1991. xxii, 368 p. Cloth $35.00.

BAYLOR, MICHAEL G., ed. and trans.: *The Radical Reformation*. New York: Cambridge, 1992. xxxvii, 295 p. Cloth $44.50, paper $14.95.

BLONDEL, ERIC: *Nietzsche: The Body and Culture: Philosophy as a Philosophical Geneology.* SEAN HAND, trans. Stanford· University Press, 1991. viii, 353 p. Cloth $39.50, paper $14.95.

BOURDIEU, PIERRE, and ALAIN DARBEL: *The Love of Art: Euoropean Art Museums and Their Public.* CAROLINE BEATTIE and NICK MERRIMAN, trans. Stanford: University Press, 1991. viii, 182 p. Cloth $29.50.

BOURDIEU, PIERRE: *Photography: A Middle-Brow Art.* SHAUN WHITESIDE, trans. Stanford: University Press, 1991 ix, 218 p Cloth $25.00.

CASTORIADIS, CORNELIUS: *Philosophy, Politics, Autonomy: Essays in Political Philosophy.* DAVID AMES CURTIS, ed. and trans. New York: Oxford, 1991. x, 304 p. Cloth $35.00, paper $14.95.

CUSA, NICHOLAS OF: *The Catholic Concordance.* PAUL E. SIGMUND, ed. and trans. New York: Cambridge, 1992. xlvii, 326 p. Cloth $69.95.

DERRIDA, JACQUES. *Acts of Literature.* DEREK ATTRIDGE, ed. and trans. New York: Routledge, 1992. xiii, 456 p. Cloth $49.95, paper $16.95

DESCARTES, RENE: *The Philosophical Writings of Descartes, Vol. III.* JOHN COTTINGHAM et al., trans. New York: Cambridge, 1991. xviii, 412 p. Cloth $65 00, paper $19.95.

ERIBON, DIDIER: *Michel Foucault.* BETSY WING, trans. Cambridge: Harvard, 1991. xiii, 374 p Cloth $27.95.

GADAMER, HANS-GEORG: *Plato's Dialectical Ethics: Phenomenological Interpretations Relating to the* Philebus. ROBERT M. WALLACE, trans. New Haven: Yale, 1991. xxx, 240 p. Cloth $27.50.

GALEN: *On the Therapeutic Method, Books I & II.* R. J. HANKINSON, trans New York: Oxford, 1991. xxxix, 269 p Cloth $72.00.

HEGEL, G. W. F.: *Elements of the Philosophy of Right.* ALLEN W. WOOD, ed.; H. B. NISBET, trans. New York: Cambridge, 1992. lii, 514 p. Cloth $59.95, paper $14.95.

HEGEL, G. W. F.: *The Encyclopaedia Logic.* T. F. GERAETS et al., trans. Indianapolis: Hackett, 1991. xlviii, 381 p. Cloth $34.95, paper $14.95.

HERACLITUS: *Fragments: A Text and Translation with a Commentary.* T M. ROBINSON, trans. Toronto: University Press, 1991. xi, 214 p. $19.95.

HONNETH, AXEL: *The Critique of Power: Reflective Stages in a Critical Social Theory.* KENNETH BAYNES, trans. Cambridge: MIT, 1991. xxxii, 340 p. Cloth $35.00.

HOPFL, HARRO, ed. and trans.: *Luther and Calvin: On Secular Authority.* New York: Cambridge, 1991. xlvi, 95 p. Cloth $29 95, paper $9.95.

KANT, IMMANUEL: *The Metaphysics of Morals.* MARY GREGOR, trans. New York: Cambridge, 1992. xii, 307 p. Cloth $49 50, paper $16.95.

KIERKEGAARD, SØREN: *Practice in Christianity.* HOWARD V. HONG and EDNA H. HONG, eds. and trans. Princeton: University Press, 1991. xix, 416 p. Cloth $49.50, paper $14.95.

KRISTEVA, JULIA: *Language: The Unknown: An Initiation into Linguistics.* ANNE M. MENKE, trans. New York: Columbia, 1991. 366 p Paper $15.00.

LE DOEUFF, MICHELE: *Hipparchia's Choice: An Essay Concerning Women, Philosophy, Etc.* TRISTA SELOUS, trans. Cambridge: Blackwell, 1991. xiii, 364 p. Cloth $49.95, paper $19.95.

LOWITH, KARL: *From Hegel to Nietzsche: The Revolution in Nineteenth-century Thought.* DAVID E. GREEN, trans. New York: Columbia, 1991. xvii, 464 p. Cloth $60.00, paper $16.50.

MARION, JEAN-LUC: *God without Being.* THOMAS A. CARLSON, trans New York: Oxford, 1992. xxv, 258 p. Cloth $32.00.

MARQUAND, ODO: *In Defense of the Accidental: Philosophical Studies.* ROBERT M. WALLACE, trans. New York: Oxford, 1991 136 p. Cloth $32.50, paper $12 95.

MERLEAU-PONTY, MAURICE: *Texts and Dialogues.* HUGH J. SILVERMAN, and JAMES BARRY, eds., MICHAEL B. SMITH et al., trans. Atlantic Highlands, NJ: Humanities, 1991. xxi, 207 p. Cloth $45.00.

MILTON, JOHN: *Political Writings.* MARTIN DZELZAINIS, ed CLAIRE GRUZELIER, trans. New York: Cambridge, 1991. xxxiii, 279 p. Cloth $44.50, paper $14.95.

POTIER, JEAN-PIERRE: *Piero Sraffa: Unorthodox Economist: A Biographical Essay.* BERNARD HOEPFFNER, trans. New York: Routledge, 1991. x, 109 p. Cloth $32.50.

PUFENDORF, SAMUEL: *On the Duty of Man and Citizen.* JAMES TULLY, ed.; MICHAEL SILVERTHORNE, trans. New York: Cambridge, 1991. xliv, 183 p. Cloth $44.50, paper $14.95.

RICOEUR, PAUL: *From Text to Action: Essays in Hermeneutics, Vol. II.* KATHLEEN BLAMEY and JOHN B. THOMSON, trans. Evanston: Northwestern, 1991. xvi, 346 p. Cloth $42.95, paper $17.95.

SAFRANSKI, RÜDIGER: *Schopenhauer and the Wild Years of Philosophy.* EWALD OSERS, trans. Cambridge: Harvard, 1991. vii, 385 p. Paper $14.95.

SARTRE, JEAN-PAUL: *Mallarmé or the Poet of Nothingness.* ERNEST STURM, trans. University Park: Penn State UP, 1991. 173 p. Cloth $25.00, paper $12.95.

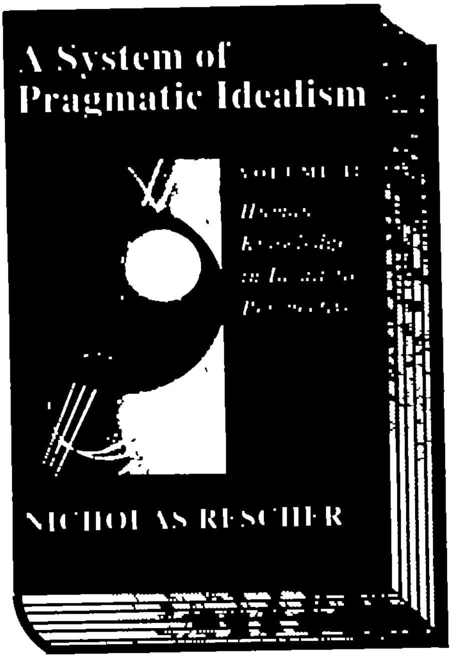
SCHLEGEL, FRIEDRICH: *Philosophical Fragments.* PETER FIRCHOW, trans. Minneapolis: Minnesota UP, 1991. xxxii, 112 p. Paper $11.95.

SIMMEL, GEORG: *Schopenhauer and Nietzsche.* HELMUT LOISKANDL, DEENA WEINSTEIN, and MICHAEL WEINSTEIN, trans. Chicago: Illinois UP, 1991. lv, 186 p. Paper $15 95.

TAMINIAUX, JACQUES: *Heidegger and the Project of Fundamental Ontology.* MICHAEL GENDRE, ed. and trans. Albany: SUNY, 1991. xxiii, 233 p. Cloth $24 50, paper $16.95.

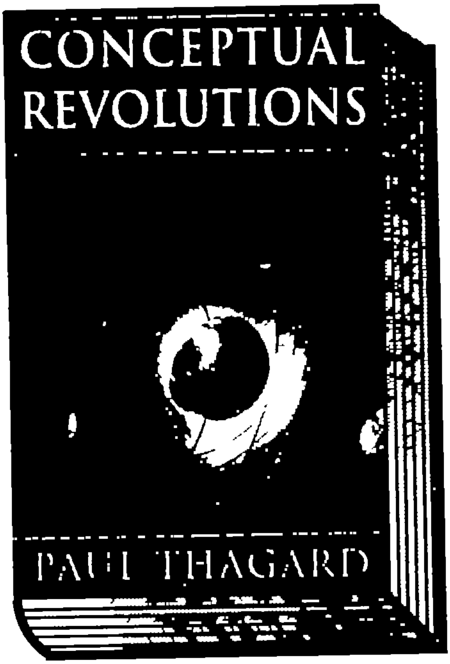
WEBER, SAMUEL: *Return to Freud: Jacques Lacan's Dislocation of Psychoanalysis.* MICHAEL LEVINE, trans. New York: Cambridge, 1991. xxii, 184 p. Cloth $47.50, paper $13.95.

WELLMER, ALBRECHT: *The Persistence of Modernity: Essays on Aesthetics, Ethics, and Postmodernism.* DAVID MIDGLEY, trans. Cambridge: MIT, 1991. ix, 266 p. Cloth $24.95.



With *Human Knowledge in Idealistic Perspective* the philosopher Nicholas Rescher launches a three-volume series that will synthesize his life's work. Rescher's numerous books and articles, which address almost every major philosophical topic, reflect a unified approach: the combination of pragmatism and idealism characteristic of his thinking throughout his career. The series will present Rescher's system as a whole, showing that rather than simply juxtaposing European continental idealism and American pragmatism in a new way, he has created an integrated philosophical position in which the central concepts of these two traditions become a coherent totality.

Cloth: $37.50 ISBN 0-691-07391-0



In this path-breaking work, Paul Thagard draws on history and philosophy of science, cognitive psychology, and the field of artificial intelligence to develop a theory of conceptual change capable of accounting for all major scientific revolutions. The history of science contains dramatic episodes of revolutionary change in which whole systems of concepts have been replaced by new systems. Thagard provides a new and comprehensive perspective on the transformation of scientific conceptual systems.

Conceptual Revolutions is essential reading for philosophers of science and is accessible to anyone interested in problems of scientific change.

Cloth: $35.00 ISBN 0-691-08745-8

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

NEW FROM PENN STATE

THE FATE OF ART

Aesthetic Alienation from Kant to Derrida and Adorno

J. M. BERNSTEIN

"Bernstein's rich and provocative study examines the essentially modern attempt to distinguish a unique or autonomous realm of the aesthetic, and presents an ambitious argument designed to undermine that post-Kantian insistence on a categorical distinction among the beautiful, the true, and the good. In doing so, he offers a thoughtful account of why the fate of art has been so central to those thinkers in the European tradition worried about the implications of the European Enlightenment, and he presents a number of original, critical readings of individual thinkers This is an important, very interesting book."

—Robert B. Pippin, University of California, San Diego

310 pages
cloth: $45.00 paper: $14.95
Literature and Philosophy Series

THE PRIORITY OF PRUDENCE

Virtue and Natural Law in Thomas Aquinas and the Implications for Modern Ethics

DANIEL MARK NELSON

This controversial interpretation of Aquinas's ethical theories serves to advance the debate about the contemporary relevance of an ethics of virtue by way of its significantly more detailed explication of prudence Nelson makes important connections between influential reinterpretations of the ethical theory of Aquinas that have been published during the last thirty years and widespread interest in an ethics of virtue that has been expressed by Alasdair MacIntyre, Stanley Hauerwas, William Sullivan, Robert Bellah, and others.

224 pages $28.50

THE PERFECTION OF THE UNIVERSE ACCORDING TO AQUINAS

A Teleological Cosmology

OLIVA BLANCHETTE

"This is a comprehensive and in-depth account of Thomas Aquinas's philosophy of nature from the perspective of the perfection of the universe. It is an important study, for it helps to fill a gap in the somewhat neglected area of Thomistic philosophy of the physical universe."

—Armand Maurer, Pontifical Institute of Mediaeval Studies

328 pages $35.00

New in paper

KIERKEGAARD'S CRITIQUE OF REASON AND SOCIETY

MEROLD WESTPHAL

"Westphal shows us that Kierkegaard's philosophy makes an important contribution to what we now call the 'critique of ideology,' embracing both political and sociological concerns, and squarely based upon an affirmation of human reason—a reason that is fully aware of its own nature, neither shirking its responsibilities nor overstepping its capacities."—Stephen N. Dunning, University of Pennsylvania

144 pages new in paper: $12.95

PLATO'S PARMENIDES

The Conversion of the Soul

MITCHELL H. MILLER, JR.

"Miller's work is a splendid achievement—original, thorough, clear, deep and exciting. He blends literary sensitivity with analytical precision in a way that helps with just about everything in one of Plato's most obscure and difficult dialogues."

—Edward N. Lee, University of California, San Diego

314 pages paper, $14.95

PENN STATE PRESS
Barbara Bldg., Suite C University Park, PA 16802 (814)865-1327

The Machiavellian Cosmos

Anthony J. Parel

In this highly original interpretation of Machiavelli's thought, Anthony J. Parel identifies a theme generally neglected in the scholarship on this sixteenth-century political thinker: Machiavelli's belief in the occult forces of heaven and humors. Challenging the current tendency to view Machiavelli as a pioneer of modern political science, Parel argues that a premodern cosmology and anthropology underlie Machiavelli's political works.

"A refreshing plea to rediscover the famous Renaissance thinker, often disfigured by the attempts to fit him in a modern frame."—Sergio Bertelli
$30.00

Now available in paperback

The Mediating Self

Mead, Sartre, and Self-Determination

Mitchell Aboulafia

In this pathbreaking book Mitchell Aboulafia considers the development of the sense of self by critically analyzing the philosophies of George Herbert Mead and Jean-Paul Sartre. Building on their work, Aboulafia provides an original analysis of consciousness and self-determination.

"[This book's] critical discussions are valuable, its original argument interesting, and its contribution to its main thesis important."—*Ethics* $11.00

Yale University Press

92A Yale Sta., New Haven, CT 06520

Now available in paperback

Justice and Modern Moral Philosophy

Jeffrey Reiman

"[Arguing for] two summary principles, that people owe each other as much noninterference as possible, and that among a co-operating group goods should be distributed according to the Rawlsian difference principle, Reiman builds a social contract theory of justice."—*Hastings Center Report* $16.00

An Inquiry into the Good

Kitarō Nishida
Translated by Masao Abe
and Christopher Ives

"A welcome new translation of a work by probably the most original and influential of modern Japanese philosophers."—Hidé Ishiguro, *Times Literary Supplement*

"Undoubtedly the most important work for anyone in the West interested in understanding modern Japanese thought."—William R. La Fleur $12.00

Understanding Peasant China

Case Studies in the Philosophy of Social Science

Daniel Little

"Students of agrarian structures and movements and rural development in China, Asia, and even elsewhere are fortunate that Daniel Little chose their bailiwick in which to exercise his skills as a philosopher of social science....The arguments are detailed and illuminating but also appropriately balanced."—Marc Blecher, *American Political Science Review*
$18.00

New from MIT

POSTMETAPHYSICAL THINKING
Philosophical Essays
Jürgen Habermas
translated by William Mark Hohengarten

This collection of Habermas's recent essays on philosophical topics continues the analysis begun in *The Philosophical Discourse of Modernity.*

Studies in Contemporary German Social Thought
224 pp $22.50 (June)

HABERMAS AND THE PUBLIC SPHERE
edited by Craig Calhoun

In this book, scholars from a wide range of disciplines respond to Habermas's most directly relevant work, *The Structural Transformation of the Public Sphere*

Studies in Contemporary German Social Thought 472 pp
$45.00

NEW PHILOSOPHY OF SOCIAL SCIENCE
Problems of Indeterminacy
James Bohman

"This book is without equal in reflecting the actual state of debate in the various social sciences.... The wide range of issues considered and the reliance on case studies from the actual practice of social inquiry make this the best available text on the new philosophy of the social sciences." — Thomas McCarthy, Professor of Philosophy, Northwestern University

280 pp $32.50

Original in Paperback
A USER'S GUIDE TO *CAPITALISM AND SCHIZOPHRENIA*
Deviations from Deleuze and Guattari
Brian Massumi

A User's Guide to Capitalism and Schizophrenia is a playful and emphatically practical elaboration of the major collaborative work of the French philosophers Gilles Deleuze and Félix Guattari.

216 pp $12.95 paper

Original in Paperback
THE NEW CONSTELLATION
The Ethical-Political Horizons of Modernity/Postmodernity
Richard J. Bernstein

"Bernstein's discussions of Foucault, Derrida, and Heidegger are models of careful commentary and fair-minded criticism." — Richard Rorty, University of Virginia

360 pp $15.95 paper

PHILOSOPHICAL INTERVENTIONS IN THE UNFINISHED PROJECT OF ENLIGHTENMENT
edited by Axel Honneth, Thomas McCarthy, Claus Offe, and Albrecht Wellmer

These 11 essays take up the philosophical aspects of Jürgen Habermas's unfinished project of reconstructing enlightenment rationality

Studies in Contemporary German Social Thought
368 pp $17.95 paper

Now Available in Paperback
AGAINST ARCHITECTURE
The Writings of Georges Bataille
Denis Hollier

An OCTOBER Book 201 pp $14.95 paper

ZONE BOOKS

Original in Paperback
WAR IN THE AGE OF INTELLIGENT MACHINES
Manuel DeLanda

A Swerve Edition 272 pp., 25 illus. $16.95 paper

Now Available in Paperback
THEORY OF RELIGION
Georges Bataille

126 pp $11.95 paper (June)

EXPRESSIONISM IN PHILOSOPHY
Spinoza
Gilles Deleuze

445 pp $16.95 paper
Distributed for Zone Books

To order call toll-free 1-800-356-0343 Fax orders (617) 625-6660 MasterCard & VISA

55 Hayward Street **THE MIT PRESS** Cambridge, MA 02142

New from MIT

THE COMPUTATIONAL BRAIN

Patricia S. Churchland and Terrence J. Sejnowski

Churchland and Sejnowski address the foundational ideas of the emerging field of computational neuroscience, examine a diverse range of neural network models, and consider future directions of the field. *The Computational Brain* is the first unified and broadly accessible book to bring together computational concepts and behavioral data within a neurobiological framework.

Computational Neuroscience Series
A Bradford Book 576 pp., 366 illus. $39.95 (May)

LIAISONS

Philosophy Meets the Cognitive and Social Sciences

Alvin I. Goldman

These essays by a major epistemologist reconfigure philosophical projects across a wide spectrum, from mind to metaphysics, from epistemology to social power.

A Bradford Book 352 pp. $39.95

BAYES OR BUST?

A Critical Examination of Bayesian Confirmation Theory

John Earman

"John Earman's *Bayes or Bust?* is a fine analysis of many issues facing modern theoretical statistics and the enterprise of confirmation theory. It will undoubtedly become a standard work among philosophers of science." — Clark Glymour, Carnegie Mellon University

A Bradford Book 264 pp. $30.00

TALK ABOUT BELIEFS

Mark Crimmins

Talk about Beliefs presents solutions to foundational problems in the philosophies of language and mind.

A Bradford Book 224 pp. $25.00

Original in Paperback

THE PHILOSOPHY OF MIND

Classical Problems/Contemporary Issues

edited by Brian Beakley and Peter Ludlow

This anthology allows readers to follow the development of thinking in five broad problem areas — the mind/body problem, mental causation, associationism/connectionism, mental imagery, and innate ideas

A Bradford Book 464 pp. $19.95 paper (June)

THE UNITS OF EVOLUTION

Essays on the Nature of Species

edited by Marc Ereshefsky

A Bradford Book 424 pp. $27.50 paper

Now in Paperback

WORLD ENOUGH AND SPACE-TIME

Absolute versus Rational Theories of Space and Time

John Earman

"This is a book about two related questions. Is motion absolute or relative? Is space a substance?.. If you are fond of a good discussion and have interest in the fundamental concepts of space, time, and motion, then this is a book for you." — *Classical and Quantum Gravity*

A Bradford Book 256 pp., 17 illus. $14.95 paper (May)

To order call toll-free 1-800-356-0343 Fax orders (617) 625-6660 MasterCard & VISA

55 Hayward Street **THE MIT PRESS** Cambridge, MA 02142

Oxford

Creation and Abortion

A Study in Moral and Legal Philosophy

F.M. KAMM, *New York University*

"Kamm is probably the most sophisticated deontologist writing on normative issues today....A major contribution to the literature."—Shelly Kagan, *University of Illinois at Chicago*

1992 192 pp. paper $9.95 cloth $27.50

Testimony

C.A.J. COADY, *University of Melbourne*

Coady explores the nature of testimony in order to show how it might be justified as a source of knowledge, and uses the insights that he has developed to challenge certain widespread assumptions in the areas of history, law, mathematics, and psychology.

May 1992 376 pp. $82.00

Natural Law Theory

Contemporary Essays

ROBERT P. GEORGE, *Princeton University*

Presenting twelve original essays by leading natural law theorists and their critics, the contributors in this book discuss natural law theories of morality, law and legal reasoning, politics, and the rule of law.

1992 384 pp. $39.95

✦*New in paperback!*

The Concept of Identity

ELI HIRSCH, *Brandeis University*

"Hirsch's book is required reading for anyone with a serious interest in identity in particular and metaphysics in general. The book is tightly argued, clearly written, and filled with fascinating material."—*The Philosophical Review*

1982 (paper 1992) 336 pp.
paper $16.95 cloth $34.50

Reason and Tradition in Indian Thought

An Essay on the Nature of Indian Philosophical Thinking

JITENDRA NATH MOHANTY, *Temple University*

Developing a new interpretation of the nature of Indian philosophical thinking, Mohanty brings to bear insights from modern western analytical and phenomenological philosophies, not so much for comparative purposes, but rather to interpret Indian thinking and to highlight its distinctive features.

1992 320 pp. $75.00

✦*New in paperback!*

Love's Knowledge

Essays on Philosophy and Literature

MARTHA C. NUSSBAUM, *Brown University*

"To move with authority over so wide a range of intellectual history, the author must be an unlikely combination: an acute and sensitive critic of ancient and modern literature, a professional philosopher and a trained scholar of ancient Greek. In this case skepticism can be dispensed with; Martha Nussbaum is all of these things."—*Washington Post Book World*

1990 (paper 1992) 432 pp.
paper $16.95 cloth $42.50

Revelation

From Metaphor to Analogy

RICHARD SWINBURNE, *Oxford University*

In this important new book by one of the world's leading philosophers of religion, Swinburne investigates the general philosophical issue of how general truth can be conveyed in unfamiliar genres, considers the general question of what would show that some book or creed conveys revealed truth, and explores what would show that the Christian creeds and the Bible convey revealed truth.

1992 248 pp. paper $19.95 cloth $65.00

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MO
To order by phone using major credit cards please call (212)679-7300, ext. 7106

Oxford University Press
200 Madison Avenue ✦ New York, NY 10016

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 5
MAY 1992

page
223 *Pointers to Truth* Haim Gaifman
262 *Is It Reasonable to Regret Things One Did?* Rüdiger Bittner
274 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1992 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY
VOLUME LXXXIX, NO 5, MAY 1992

POINTERS TO TRUTH*

Consider the following:

Jack: What I am saying at this very moment is nonsense
Jill: Yes, what you have just said is nonsense.

Apparently, Jack spoke nonsense and Jill spoke to the point. But Jack and Jill seem to have asserted the same thing: that Jack spoke nonsense. Wherefore the difference?

To avoid the vagueness and the conflicting intuitions that go with 'nonsense', let us replace it by 'not true' and recast the puzzle as follows:

line 1: The sentence on line 1 is not true.
line 2: The sentence on line 1 is not true.

If we try to evaluate the sentence on line 1 we find ourselves going in an unending cycle. For this reason alone we may conclude that it is not a true sentence. Moreover, we are driven to this conclusion by an elementary reasoning: if the sentence on line 1 is true then what it asserts is true, but what it asserts is that the sentence on line 1 is *not* true. Consequently, that sentence is not true. But when we write this true conclusion on line 2 we find ourselves repeating the very same sentence. It seems that we are unable to deny the truth of the sentence on line 1 without asserting it at the same time.

---

* The paper's basic ideas were presented in October 1985, at the Bar-Hillel memorial conference held at Boston University. Further developments were reported in various lectures and invited talks at: Harvard, Princeton, Stanford, UCLA, UC Irvine, CSLI, the Pacific APA meeting, the University of Texas/Austin, and the 1991 ASL meeting. The technical core was presented in the 1988 TARK meeting and published in the proceedings (cf. fn. 3). I have benefited from many reactions. From the many to whom thanks are due I shall only mention David Kaplan, Rohit Parikh (who found a bug in one of the earlier proofs), Charles Parsons, Hilary Putnam, and Brian Skyrms.

0022-362X/92/8905/223–61

Essentially, the puzzle is a variant of what Bas van Fraassen[1] has named "the strengthened liar"—called by some "the metaliar"—which I shall simplify here to *the strong liar.* We might recall at this point that the usual version of the liar paradox consists in the fact that the sentence on line 1, or some analog of it, is neither true nor false, because either assumption leads to contradiction. That version can be resolved by admitting gaps, i.e., by giving up the rule that every sentence must be either true or false. The present paradox is an altogether different kettle of fish. It consists in our seeming inability to express a true conclusion—namely, that a certain sentence is not true—without repeating the very same sentence whose truth we deny, or one equivalent to it.

It is known that the semantic paradoxes can be reconstructed in various modal frameworks (cf. Richard Montague[2]). Indeed, using a knowledge predicate we get an analogous puzzle:

line 1: What is written on line 1 is not known by Jill to be true.

Regarding the sentence, Jill concludes that, indeed, she does not know it to be true. (If she knows it to be true, then it is true; but, given what the sentence says, this implies that she does *not* know it to be true, hence—a contradiction.) Jill writes her conclusion on line 2:

line 2: What is written on line 1 is not known by Jill to be true.

So Jill knows the truth of what is written on line 2, because she has just deduced it, but not of what is written on line 1. A similar puzzle can be made with 'known' replaced by 'believed'. With 'necessary' we get the following version:

line 1: The sentence on line 1 is not necessarily true.

line 2: The sentence on line 1 is not necessarily true.

Again, the line-1 sentence is not necessarily true; because if it were, it would *not* be true and, a fortiori, not necessarily true. This conclusion *is* necessarily true, because we have just proved it. Writing it on line 2, we get a necessarily true sentence, which merely repeats the not necessarily true sentence on line 1.

The moral of all these puzzles is simple. In situations of this nature we should assign truth values not to sentence types but to their tokens. Having concluded that the line-1 sentence is not true, we state the conclusion by displaying another token of the very same sentence. The second token, on line 2, expresses something altogether different from what is expressed, if anything, by the token on line 1. The people who agreed with Jill in the Jack and Jill exchange had no difficulty in grasping this point.

---

[1] "Presumption, Implication, and Self-reference," this JOURNAL, LXV, 5 (March 7, 1968): 136–52.

[2] "Syntactical Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability," *Acta Philosophica Fennica*, XVI (1963): 153–67.

In this respect, modal predicates like 'know', 'believe', and 'necessary' are in the same boat with 'true' and the same remedy is required, namely, to allow different tokens of the same sentence to have different meanings, or—if you want—to express different statements, or different propositions. One should expect the formalisms to differ according to the predicate in question, but the same general framework will underlie them. In the present work, I set up the formalism for truth, thereby providing also the basic framework for the various modalities.

Following the intuition just given, we can see that in general what a token expresses depends on (1) what it *says*, i.e., on the sentence type and (2) on the whole network: on the tokens to which the sentence refers and on the tokens to which *they* in their turn refer, etc. This is what distinguishes the self-referential sentence token on line 1 from its non-self-referential brother on line 2.

I shall propose a formal setup for specifying networks of tokens and an algorithm for assigning truth values. It assigns the sentence token on line 1 the value *GAP*, that on line 2 the value *T* (True), and it yields the same kind of intuitive results in general.

The following "Preview" contains a rather nontechnical presentation of the method, with an illustrative example. It also contains philosophical points and clarifications concerning the proposal's significance, some of which are elaborated in later sections. In section II—where the concept of a ***black hole*** is defined—a wider perspective is given, which clarifies the proposal's place and motives, within the frameworks of natural-language semantics, and in the philosophy of mathematics. Section III is devoted to the issues of indexicality, levels of truth, and the import of a non-Tarskian semantics. Section IV discusses some points concerning pointers and propositions, and comments on some aspects of the Barwise-Etchemendy work. In section V, previous accounts of the strong liar are discussed: those of Charles Parsons, Tyler Burge, and Brian Skyrms.

I PREVIEW

I base the formalism on three truth values *T*, *F*, and *GAP*. I shall refer to the first two as the ***standard truth values***. The third value, *GAP*, signifies more than mere absence of a standard value. It signifies ***recognized failure***. The idea is that we assign tokens, or—as we shall later see—pointers, the value *GAP* when, in the course of the evaluation, we conclude that they must fail, i.e., that the rules will not enable us to assign them a standard value. We can then go on and use other, "uninfected" tokens to assert of a previously failed token that it is not true, or not false, or that it is neither true nor false.

Those acquainted with Kleene's three-valued logic will be quick to match *GAP* with Kleene's *u*, which stands for 'undefined'. *GAP* is, however, an active value, not a mere "undefined"; an assignment of it can serve as a basis for further assignments of *T*, or of *F*. By making *GAP* an active value we can construct on top of the gap instead falling into it.

Now, the concept of a token is too narrow for the purpose of a general framework. For we might want to refer to sentences without having them displayed somewhere as tokens. Suppose that there is just one sentence on line *n*, then we can use 'the negation of the sentence on line *n*' in order to refer to the negation of that sentence, independently of there being a token of that negation. And, in general, we might speak about the sentence obtained from that on line *n* by any syntactic manipulation, e.g., the substitution of some term by another. Therefore, I use a more general concept, that of a pointer. *A pointer is any object that is used to point to some object.* A token constitutes a special kind of pointer, *it points to the type of which it is a token.*

Pointing is determined by a variety of conventions. Being a pointer is not something intrinsic in an object, but amounts solely to its functioning in a certain role. Any object can serve. If desired, we can regard descriptions as pointers to the described objects. And one can set up a formalism whereby every sentence type is a pointer to itself. The solution to the semantic paradoxes requires, however, that we have many pointers to the same sentence type, so that when a pointer fails another is available. Any object can be pointed to. But since in this paper only pointers to sentence types are needed, I shall reserve 'pointer' for this case, unless the context indicates otherwise.

I find it is useful not to enter, at least at this stage, into the nature of pointers, but to take them as primitives. Thus, I assume a language, a set of objects called "pointers," which have names in that language, and some fixed arbitrary function that correlates with each pointer a sentence (of that very same language) to which it points. A pointer whose name occurs in some sentence can point to any sentence whatsoever (including the sentence containing its name). All possibilities of direct or indirect self-reference are thus covered.

Only the following additional structural elements are needed for the system. We have to correlate, with each pointer to a compound sentence, pointers that point to its sentential components. Thus, for any pointer p pointing to $A*B$, where $*$ is a sentential connective, there are pointers $p1$ and $p2$, which point, respectively, to A and to B. And if p points to the negation $\neg A$, then $p1$ points to A (and we can, in this case, stipulate that $p2 = p1$). I call $p1$ and $p2$ the *derived pointers* of p. If the language has quantifiers, we also correlate with every p that points to $\forall v A(v)$, or to $\exists v A(v)$, and with every constant term t, a derived pointer $p|t$, which points to $A(t)$ (the result of substituting t for v in $A(v)$).

Note that, if p is a token, then $p1$ and $p2$ can be simply identified with the corresponding subtokens. But with quantifiers, we must go beyond tokens, for in general we cannot identify $p|t$ with a token. For our present purposes, as long as quantifiers are not included, we can take our pointers to be just tokens. The reader may, for the sake of convenience, assume this to be the case. The extension to the general case is straightforward.

The upshot of this approach is a new kind of semantics in which truth values are assigned to pointers and the usual recursive definition of truth is replaced by a set of rules for evaluating networks. Here is an informal description of the algorithm, illustrated by a simple example. This, I think, should give an adequate idea of the method, providing thereby a basis for the discussion that follows. (A full mathematical presentation is planned for a separate companion paper. A concise version containing the formal definitions and the statements, without proofs, of the main theorems can be found in my TARK paper.[3])

Consider the above mentioned system of language and pointers. Assume that the language (based on predicates and, possibly, function symbols) is interpreted, except for two, so-called *truth predicates* (or *semantic predicates*): $Tr(\ )$ for 'true' and $Fa(\ )$ for 'false'. These two predicates take pointer names as arguments. Assuming that p, q, r, etc., are pointers, and that the pointer names used in this discussion are also their names in the formal language, we get sentences of the form $Tr(p)$, $Fa(q)$, etc., where p, q, etc., are pointers to sentences of this very same language.

---

[3] "Operational Pointer Semantics: Solution to Self-referential Puzzles I," *Theoretical Aspects of Reasoning about Knowledge* [hereafter TARK], M. Vardi, ed. (Los Angeles: Morgan Kauffman, 1988), pp. 43–59.

The goal is an assignment of truth values to all the pointers. It is achieved by an ***algorithm***—also called the ***evaluation procedure***—which is based on a collection of rules for assigning truth values. They are naturally divided into three groups. The first, the ***standard rules*** group, consists of rules that correspond to the traditional rules for assigning standard values; they include also rules for the truth predicates.

> *Standard rules.* Pointers to sentences of the form $P(t_1, \ldots, t_n)$, where P is a nonsemantic predicate, get their truth values in the usual way, via the interpretation of the predicates and the denoting terms. Such sentences are called *basic*. Evidently, different pointers to the same basic sentence get the same truth value. It is a standard value determined by sentence type alone. And this holds as long as only standard rules are employed.
>
> Now, if p points to a nonbasic sentence, then, in the usual order of things, its value is determined by the values of other pointers, which, so to speak, are *called directly by p*. Thus, if p points to the disjunction $A \vee B$, then the pointers it calls directly are its derived pointers $p1$ and $p2$ (which, recall, point to A and to B). If one of them gets the value T, so does p (no matter whether the other has been assigned a value, or what that value is). If both get F, p gets F as well. The standard rule for negation, i.e., the "toggling" of the standard truth value, should be obvious. Other connectives are treated by expressing them in terms of $\vee$ and $\neg$.
>
> A pointer, p, to a quantified sentence, calls directly each of the pointers $p|t$ to the substitution instances. The rules for quantified sentences are obtained by treating them as (possibly infinite) disjunctions or conjunctions. For the sake of convenience, I assume that every object in the interpretation's universe is named by some term and that quantifiers are interpreted substitutionally. This is not essential. Quantifiers can be interpreted referentially, at the cost of adding some obvious items.[4]
>
> Finally, a pointer to $Tr(q)$, or to $Fa(q)$, calls directly q, if q gets a standard value, then p gets, in the first case, the same value; in the second case, the opposite value. These are the standard rules for the truth predicates.

The essential part of the algorithm, by which it enhances classical logic, is constituted by two additional groups: the ***gap rules*** determine the cases of failure, where GAP is assigned. The ***jump rules*** determine assignments of standard values, which are based on previous failures. They enable us to use new pointers in order to make semantic assertions about previously failed ones. Before I go into these two groups, here is a simple illustrative example.

---

[4] Namely, pointers would point to well-formed formulas (not only to sentences) and truth values would be given to pairs consisting of pointers and assignments of objects to variables

Let John, Joan, Jack, and Jill make the following statements:

John: What Jill says is not true and what Joan says is not true.
Jill: What Jack says is true.
Joan: What Jill says is not true.
Jack: Either McX's conjecture is true or what Jill says is false

For simplicity, we can use here each speaker as a pointer to his uttered sentence. The resulting network is represented in figure 1. The arrows between pointers (the nodes) do not represent pointing, which is a relation between pointers and sentence types, but the above mentioned direct calls between pointers. McX's conjecture, represented as a box, is not spelled out. It might involve a whole network of its own; but I assume, for the purpose of illustration, that it does not refer, directly or indirectly, to any of the four utterances.

Actually, the diagram has been simplified by merging certain arrows and hiding certain derived pointers; e.g., the arrow from John to Jill should be split into an arrow from John to John1 (which points to the first conjunct of what John says), an arrow from John1 to John11 (which points to 'What Jill says is true'), and an arrow from John11 to Jill. Note that the pointed-to sentence types can be read from the diagram, though they are not explicitly displayed.

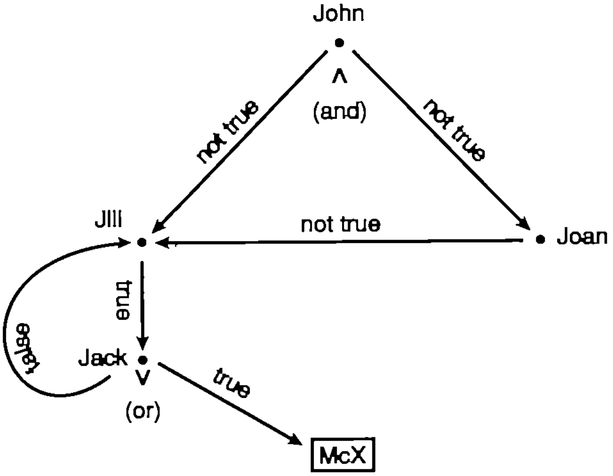


Figure 1

The algorithm is applied recursively. Assume that it results in assigning *F* to McX. This leaves Jack and Jill in a closed unresolved loop. Both get at this stage the value *GAP*. Then Joan, who makes an assertion about Jill and does not belong to the loop, gets a standard value (here a jump rule is employed). Since her assertion is true, she gets *T*. In the same way, the first conjunct in John's assertion gets *T*. But the second conjunct (which asserts the nontruth of Joan) gets *F*, hence John gets *F*. Had McX got *T*, Jack and Jill would have got *T*, while John and Joan would have got *F*.

In the case of the two-line puzzle the network is:

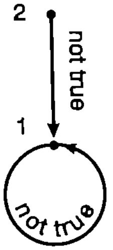


Figure 2

For '*i*' read 'the sentence token on line '*i*'. Right at the beginning we get a closed loop consisting of the pointer 1. It gets the value *GAP*. Then 2 gets the value *T*.

In the four-speaker example, the standard rules for *Tr* and for negation are used for determining the value of John2 (which points to the second conjunct of what John says). Also, the standard rule for conjunctions is used for determining the value of John. The additional rule groups are as follows.

> *Gap rules*. The most important of these is the *closed loop rule*, which states the following: if, in the course of applying the evaluation procedure, a closed unevaluated loop forms and none of its members can be assigned a standard value by any of the rules, then *all* of its members are assigned *GAP* in a single evaluation step.
> [A *closed unevaluated loop* is a set, say *L*, of unevaluated pointers (i.e., which have not been assigned truth values), satisfying the following two conditions: (1) unevaluated pointers called directly by any member of *L* also belong to *L*; (2) from each member of *L* one can reach any other member of *L* by one, or more, finitely many direct calls; i.e., by a sequence $p_1, \ldots, p_n$, where $n > 1$, the p_is are in *L* and each calls directly its successor.]

In the four-speaker example, Jack and Jill get their value via the closed loop rule. And in the two-line example this rule applies to the token on line 1, which forms by itself a closed unevaluated loop. In both examples the loops are simple cycles. But, in general, the closed unevaluated loop can be as complex as you can imagine and might consist of any number, finite or infinite, of pointers.

There is also a ***simple gap rule*** that states: a pointer is assigned *GAP* if all pointers called directly by it have been assigned values, but the pointer cannot be assigned a standard value by any of the rules.[5]

> ***Jump rules.*** Assume that q points either to $Tr(p)$ or to $Fa(p)$, and that p, but not q, has already been assigned *GAP*. Then the jump rules (for Tr and for Fa) assign q the value F.

To see how the jump rules work assume that an unevaluated r points to $\neg Tr(p)$ and that p has been assigned *GAP*. Then, by the jump rule for Tr, $r1$ (which points to $Tr(p)$) gets F; and, by the standard negation rule, r gets T. Thus, the jump rules provide for the possibility of successfully asserting that a given *GAP* pointer is not true, or that it is not false. ***Jump*** signifies here true ascent in the metalinguistic hierarchy. In our example, a jump rule (for Tr) is needed in order to assign Joan and John1 their values. It is also invoked in assigning a value to the line-2 token in the two-line puzzle.

To sum up, the assignment of *GAP* signifies recognized failure: we decide that the pointers in question fail to evaluate to a standard value. The grounds for such a decision are given in the gap rules. The assignment of *GAP* can serve as a basis for further assignments of standard values via the jump rules. What is inexpressible in the usual denotational semantics is thus expressible through network evaluation.

---

[5] Given the standard rules, the effect of this rule is to enforce Kleene's strong three-valued truth tables A pointer p to a disjunction $A \vee B$ is assigned *GAP* if $p1$ and $p2$ have been assigned values that do not enable us to assign, by some other rule, a standard value to p; this is the case where one of $p1$, $p2$ is assigned *GAP* and the other is assigned either *GAP* or F. In other variants, however, based on a different set of standard rules (e.g., using supervaluations), the simple gap rule does not imply Kleene's strong truth tables.

The gap rules determine the assignments of *GAP* that signify recognized failures. But *GAP* can be assigned also by an additional so-called ***give-up*** rule: if at some stage we are left with unevaluated pointers to which none of the other rules applies, then all of them are assigned *GAP*. Gaps assigned by this rule do not constitute ***recognized*** failures, but are rather like Kripke's "undefined" u It can be shown that in all networks of finitary type, i.e , where every pointer calls (directly or indirectly) finitely many pointers, the give-up rule is never employed. The aim of more sophisticated versions of the algorithm is to avoid, as far as possible, the employment of such a rule in nonfinitary networks.

Various mathematical results, which I shall not go into here, establish the desirable properties of the evaluation procedure. I shall only describe informally two of the basic results, whose significance, I hope, can be seen:

> *Order independence*. The rules can be applied in any order, to whatever pointers; the (possibly transfinite) sequence of applications results in a unique total valuation which does not depend on the order of applications.[6]
>
> *Limited pointer dissent* (abbreviated as LPD). Two pointers to the same sentence are assigned different values only if one of the pointers is assigned *GAP*. This is also true if the pointed-to sentences are logically equivalent.

The evaluation procedure, also called an "algorithm," has indeed the form of an algorithm: an exact prescription for applying a well-defined set of rules. But this "algorithmic" form should not disguise its true potential. It applies to all networks, including infinite ones of arbitrary cardinality, and it extends both the system of Alfred Tarski and that of Saul Kripke.[7] Each of these is obtained by deleting from my algorithm certain rules: Tarski's truth definition is what we get if we employ only the standard rules for basic sentences, connectives, and quantifiers. Kripke's system is what we get if we employ only the standard rules—including the standard rules for the truth predicates. (In either case, pointers remaining at the end unevaluated are assigned *GAP*, or "undefined.") It can be shown that, if the jump rules are not employed, then a pointer's value depends only on the sentence it points to; truth values would be thus determined by sentence *types* alone.

Note that, in Kripke's construction, both tokens in the two-line puzzle constitute gaps (i.e., remain unevaluated), as do the utterances of the four speakers. In the variants of Anil Gupta[8] and Hans Herzberger[9] they are assigned oscillating values. The two major points of difference between my proposal and all the previous ones are, first, the use of pointers and, second, the gap and jump rules.

---

[6] A rule application extends a given partial valuation. If only the standard rules are used, then these extensions constitute monotone operators and the order-independence claim follows easily by a fixed-point argument. With the gap and jump rules, however, the operators are not monotone. The claim is nonetheless true, but the proof is much more involved.

[7] "Outline of a Theory of Truth," this JOURNAL, LXXII, 19 (November 6, 1975): 690–716. Repr. in *Recent Essays on Truth and the Liar Paradox*, Richard Martin, ed. (New York: Oxford, 1984). The essays cited in fns. 8, 9, 12, 27, and 64 are also reprinted in this collection.

Recovering the metalanguage hierarchy. There is a precise definition, based on the evaluation procedure, which correlates with each pointer an ordinal number expressing an intuitive notion of "metalinguistic level." The pointers can therefore be stratified, according to their levels, so as to reflect a Tarski-like hierarchy.

Roughly speaking, a pointer's level is the number of *nested applications of jump rules* needed in order to evaluate it. In the four-speaker example, if α is the level of McX, then also Jack and Jill are on level α, whereas John and Joan are on level $\alpha + 1$. We can also refine each of our levels so as to reflect a minor, or local, subhierarchy within it. Roughly speaking, the minor subhierarchy reflects the height of nested applications of the *standard rules for the truth predicates,* which are necessary—once the major level has been reached.

Unlike Tarski's hierarchy, we have a single language with a single truth predicate and the stratification is effected after the fact, i.e., after the assignment of truth values. In this respect it is like the hierarchy obtained via Kripke's construction; as he aptly puts it, the pointers (in his case—sentences) seek their own levels. But his whole hierarchy of grounded sentences corresponds (exactly) to our local subhierarchy of level 0. Real climbing starts where Kripke leaves off. All our nonzero levels are in his model gaps ("undefined").

The construing of truth as a predicate over *pointers* rather than types is related to the metalinguistic hierarchy as follows. Call pointers *dissenting* if they point to the same sentence and have different truth values. We know, by LPD, that of any two dissenting pointers one is assigned *GAP*. It turns out that the non-*GAP* pointer is one level higher in the hierarchy:

> *Higher level dissent* (abbreviated as HLD). For every two dissenting pointers, p_1 and p_2, if p_1 is the *GAP* pointer then $l(p_2) = l(p_1) + 1$, where $l(p_i)$ is the level of p_i.

The significance of LPD and HLD will be discussed in section II. Before proceeding let me clarify certain points, answering thereby possible questions and allaying, perhaps, certain misgivings that might have arisen. Some of these points will be reconsidered, at greater detail, in later sections.

---

[8] "Truth and Paradox," *Journal of Philosophical Logic*, XI (1982): 1–60.

[9] "Notes on Naive Semantics," *Journal of Philosophical Logic*, XI (1982) 61–102

What pointers express. Some might find the idea of assigning truth values to pointers strange; the phrase 'that pointer is true' may sound bizarre. Truth and falsity, so it is argued, have to do with what we *say*. Surely, it is inappropriate to endow some arbitrary objects, used for pointing to sentences, with truth or falsity. To argue thus is to misunderstand my proposal. The phrasing is not essential. It is not strange to speak of the truth or falsity of sentences, meaning thereby sentence tokens; and if it comes to pointers, one could speak of a pointer pointing to a truth, or to a falsity, or indicating, or failing to indicate one. Nothing marks a pointer as such except its function within a global arrangement. But this very function makes it a tool for expression, hence—something on which truth values can be hanged.

As far as the ontology of statements, propositions, meanings, etc., is concerned, my proposal is neutral. A nominalist who bars intensional entities can use something like the evaluation procedure as part of the account of the functioning of language. But the proposal is compatible with seeing pointers (in the contexts of networks) as means of expressing propositions, or statements.[10] On a pretheoretic level, it is not difficult to make sense of "the proposition expressed," or "the statement made" by means of a pointer. For example, what is expressed by the sentence token on line 2 is simply that the sentence token on line 1 is not true, which is to say: *is not assigned the value T by the evaluation procedure.* Although the token on line 1 is of the same sentence type, it fails, because of the loop, to express this. Similarly, in the Jack and Jill exchange, Jack says of himself that he does not express a proposition, but *in this very saying no proposition is expressed* (at least not the proposition indicated by his words, at their face value.) What Jack seems—but has failed—to express is successfully expressed by Jill. If only sentence types are considered, Jack and Jill appear to have stated the same thing. Actually they did not. The source of the illusion is that they employed sentence types that are, in the given conversational context, logically equivalent.

The following can perhaps add clarification. Two intuitive theses about truth lead to the *strong liar:*

---

[10] Cf fn. 55 concerning 'propositions' and 'statements'.

(T1) If the sentence '*x* is not true' is true, then *x* is not true.

(T2) If *x* is not true, then the sentence '*x* is not true' is true.

Here '*x*' ranges over sentence types and quotes are interpreted as Quine's semiquotes (i.e., the quoted sentence, which is *x*'s value, is to replace '*x*'). We get the strong liar when *x* is, or is equivalent to, '*x* is not true'. But if we construe 'true' as a predicate of pointers, then '*x*' should range over pointers and (T1) and (T2) should be modified by replacing 'the sentence' by 'any pointer to the sentence'. We can still maintain (T1) in its new form, but we have to qualify (T2) by rewriting it as:

(T2*) If *x* is not true then any ***unfailed*** pointer to '*x* is not true' is true.

The paradox is thus blocked by allowing pointers—in particular, tokens—to fail, and by giving up (T2) for failed tokens of '*x* is not true'. If we employ the analogous strategy for the case of types, and *x* is a liar sentence, then it is the *type* '*x* is not true' which fails. Thus, we cannot use it to say that *x* is not true. Further observations concerning pointers and propositions are made in section IV.

An essentially non-Tarskian semantics. In Tarskian semantics, truth is a property of sentence types; tokens as such are not considered. That tokens play a central role in natural language, due to the abundance of indexicals and demonstratives (words like 'I', 'you', 'now', 'this', 'that') is, of course, very old news. These terms, having highly variable context-dependent denotations, cause truth value to vary according to token.

But it would seem that the Tarskian framework can be adjusted, or rather supplemented, so as to handle these phenomena. Having replaced the items of varying denotation by others, whose denotations—within the considered discourse—are fixed, we can apply Tarskian semantics. To use an admittedly simple example, transform 'He is tall' to '*X* is tall', where '*X*' is an appropriate context-insensitive name or description of the indicated person, and evaluate the result as a sentence *type*. If needed, a further reduction can be used to eliminate the temporal indexicality of 'is'. The truth predicate that is defined over tokens is thus reduced to a predicate defined à la Tarski over types.

The present situation is altogether different. The predicate defined by the evaluation procedure is ***irreducibly over tokens,*** or—

more generally—over pointers.[11] Here, as I shall show, differences in truth value are not traceable to terms that behave like indexicals, or have ambiguous denotations, which occur in the sentences.

Attempts, like Tyler Burge's,[12] to resolve the semantic paradoxes by construing 'true' as an indexical cannot but fail to achieve that goal. Indeed, as the analysis will show, Burge's account puts all the burden of work on a rather obscure mechanism of implicatures, leaving indexicality idle. (The general argument concerning indexicality is laid out in section III; the more specific discussion of Burge's work is in section V.)

Much more significant than the failure of indexicality is the change from compositional semantics, whereby the meaning of a sentence is "composed" of the meaning of its parts, to a radically different one, in which tokens, or pointers, are evaluated globally. The present proposal shows that this essentially different picture can be embodied in a formalism as precise as Tarski's truth definition for formal languages.

An argument propounded by Donald Davidson has it that the ability of finite beings—such as humans—to use a potentially infinite number of sentences is to be explained by assuming a finite number of rules,[13] whereby meanings of compound expressions are determined by the meanings of their components. A setup of this nature resembles Tarski's truth definition. That something of this kind underlies a great deal of the functioning of language cannot be doubted. But it is equally true that a great deal lies beyond it. If human ability for controlling a potential infinity is to be modeled by a formal device—a plausible working assumption, though far from an axiom—then the appropriate concept is that of an *algorithm*. A finitely axiomatized compositional semantics amounts to a rather special kind of algorithm. Appealing as the model is, it is more than plausible that the processing of language, on the level of meaning, involves also algorithms of quite a different nature. I shall say more on this score in section III.

---

[11] Formally, one can of course redefine "type" by stipulating that tokens that get different truth values (e.g., those on lines 1 and 2) are to be considered of different type. Isaac Levi observed that such a stipulation might not be bizarre, given that the token-type distinction is not always clear-cut and that homonyms might be considered as different types, because of their different meanings. But if we do this, the very definition of "type" presupposes the assignment of truth values to tokens; in this sense, the truth predicate is still irreducibly over tokens.

[12] "Semantical Paradox," this JOURNAL, LXXVI, 4 (April 1979): 169–98.

[13] On the syntactical level, Noam Chomsky had argued in this way for an innate language-processing mechanism

Truth, falsity, and GAP. In the pointer formalism, the predicates representing truth and falsity figure as two primitives. The evaluation rules relate truth and falsity, through negation, as follows. *A pointer to the negation* ¬A gets the value T, or the value F, iff the derived pointer to A gets the opposite value. Thus, our concept of falsity accords with the well-known thesis, according to which a sentence is false iff its negation is true. It is possible to set up an equivalent formalism that uses only *Tr*, or only *Fa*, as a primitive.[14]

Let 'false$_1$' denote falsity in the sense employed here. Another way of relating falsity to truth is given by 'false$_2$'. *A sentence is false$_2$ just when it is not true.*[15] Consequently, every sentence is either true or false$_2$. When one reads 'false' as 'false$_2$' there is a miraculous disappearance of truth-value gaps. Actually, they are there, thinly disguised and easy to uncover. A sentence is a gap if and only if both it and its negation are false$_2$.

As the last equivalence suggests, easy translations are available from one 'false$_i$' terminology to the other, making the two equivalent for purposes of expression. The present formalism can be recast in terms of 'false$_2$', at some cost of additional structure.[16]

The false$_2$ interpretation might remind one of cases, arising in Bertrand Russell's theory of descriptions, where sentences and their "negations" come out false; e.g., (i) 'The present king of France is bald' and (ii) 'The present king of France is not bald'.[17] But the similarity is only apparent. The falsity of (ii) is caused by the narrow scope of the negation. True negation—one that applies to the whole sentence—*does* convert falsity to truth: (iii) 'It is not the case that the present king of France is bald' comes out true. Much of the appeal of Russell's theory is due to its ability to distinguish between local negation, as exemplified in (ii), and true negation, as exemplified in (iii). There are no gaps here. Acceptable or not, Russell's theory rests squarely within the framework of classical two-valued logic.

---

[14] We can define the evaluation procedure for a subsystem based on a single truth predicate. But to have the expressive power of our present system, we need a way to express both $Tr(p)$ and $Fa(p)$, e.g., both are needed in order to say that p is a gap $(\neg Tr(p) \wedge \neg Fa(p))$. This requires an additional structural item for pointers: a negation operant that associates with every pointer p a pointer, say $\mathit{neg}(p)$, to the negation of the sentence pointed to by p. We can then translate $Fa(p)$ into $Tr(\mathit{neg}(p))$ (We should stipulate that $\mathit{neg}(p)1 = p$ and that if p points to a negation then $\mathit{neg}(p1) = p$). Note that, in general, $\mathit{neg}(p)$ will not be a token, even if p is. The present formalism avoids the need for a negation operant; hence, unless we consider quantifiers, we can assume that all our pointers are just tokens.

[15] Such a reading of 'false', for the case of certain atomic sentences, was suggested in David Kaplan's "What is Russell's Theory of Descriptions?" *Physics, Logic, and History,* W. Yourgrau and A. D. Beck, eds. (New York: Plenum, 1970). The same reading also turns up in Jon Barwise and John Etchemendy's *The Liar: An Essay in Truth and Circularity* (New York· Oxford, 1987). But the issues there, to be discussed in section IV, are complicated by the treatment of negation.

[16] We shall need a negation operant for pointers; cf. fn. 14.

[17] I am ignoring here the more radical Russellian position according to which definite descriptions are incomplete symbols to be eliminated, in the final account, altogether.

The present situation is altogether different. The nontruth of the line-1 sentence (in the two-line puzzle) does not hinge on the restricted scope of 'not'. Making that scope as wide and as unambiguous as possible does not make the slightest difference. Consider for example:

line 1: *It is not the case that the sentence on line 1 is true.*

To classify this last sentence as false means to classify as false both a sentence *x* and 'it is not the case that *x*'.[18] And while it is conceivable for both *x* and 'it is not the case that *x*' to fail, it is inconceivable that both be false—unless 'false' is just another mark for failing. That is, 'is false' is to be read here as *is nontrue,* or—still better—*fails to be true,* or *fails to express a truth.* To what extent this subsumption of failure under falsity reflects ordinary usage is a question for the lexicographer.

A two-fold classification of all sentences into "true" and "false" may serve to make a philosophical point.[19] But it should be clear that by no means does it signify a two-valued logic. There is more to two-valued logic than there being just two "truth values." The two values should underly the semantics of the sentential connectives. That they should work well with negation is the minimum requirement.

To sum up, the construing of 'false' as 'false$_2$' has no bearing on the existence of gaps, but is merely a terminological grouping of falsity (i.e., truth of the negation) together with other kinds of failures. There is a moral in this concerning propositions. Either we say that liar-type sentences express no propositions, or we make place for propositions that constitute truth-value gaps.

The liar and the truth teller. The algorithm described so far treats the truth teller (a sentence asserting its own truth, e.g., 'What I am saying now is true') in the same way as the liar; both get *GAP*. Our framework does, however, provide means for distinguishing between the two. The distinction is made by considering consistent assignment of truth values, or what I call in the technical paper *self-supporting valuations.*[20] There is a total self-supporting valuation that assigns the truth teller (token) the value *T*, and there is one that assigns it *F*. But every self-supporting total valuation must assign the liar (token) the value *GAP*.

---

[18] Take as *x* the sentence 'the sentence on line 1 is true'. If the sentence on line 1 is false, so is *x* and so is 'it is not the case that *x*'

[19] As was done, for example, by Michael Dummett, in "Truth," *Proceedings of the Aristotelian Society,* LIX (1959): 141–62.

Different variants of the algorithm. The present variant of the algorithm leads to a truth-value assignment that conforms to Kleene's three-valued truth tables (cf. fn. 5). Some may, however, prefer to assign *T* to all pointers to tautologies. This can be achieved by using a variant in which the set of standard rules is replaced by another set, based on *supervaluations*. In the usual supervaluation method—which applies to sentence types—a sentence is given a standard value whenever it gets this value in all possible complete standard extensions of the partial valuation. This idea can be adopted to assignments of values to pointers; I defer details to the technical paper.

It is noteworthy that, to get the supervaluation variant, only the standard rules are changed. The gap and jump rules remain the same.[21] In other variants as well, changes can be done "locally" within this or that group of rules. This illustrates the modular character of these groups.

More sophisticated variants, not to be presented here, incorporate additional gap rules. These rules introduce grounds for discriminating between pointers. They enable early assignments of *GAP* to some of the pointers, thereby salvaging the remaining ones that would have otherwise been all members of the same closed loop. The total effect is to reduce the assignment of *GAP* and to make more pointers available as means of expression. Such variants aim at achieving higher degrees of self-contained systems for some highly expressive languages—a goal that will become clearer in the next section.

II BLACK HOLES

Suppose that we assign truth values to sentence types. Consider *any* notion of sentence equivalence, which preserves truth value and which satisfies the following two mild conditions. (i) Each sentence is equivalent to itself. (ii) If '. . .' and '——' are terms referring to equivalent sentences, then '. . . is not true' is equivalent to '—— is not true'.[22] Let us now extend the original two-line text as follows:

---

[20] Roughly speaking, a self-supporting total valuation is one that is closed under rule applications: applying to it the rules of the algorithm yields the same valuation.

[21] Changing the standard rules can affect the applicability of the gap rules, because a gap rule can apply only if other rules cannot. The gap rules themselves, however, need no change

line 1: The sentence on line 1 is not true.

line 2: The sentence on line 1 is not true.

line 3: The sentence on line 2 is not true.

line 4: The sentence on line 3 is not true.
⋮
line n + 1: The sentence on line n is not true.

The sentences on the first two lines are the same, hence—equivalent. And, by (ii), the equivalence of the sentences on lines $k - 1$ and k implies that of the sentences on lines k and $k + 1$. Consequently, all the sentences are equivalent. Since the sentence on line 1 is not true, the sentence on line 3 is not true either, but what I have *just stated* is not true, because it is the sentence on line 4 (or an obviously equivalent reformulation of it), and also *this* last statement of mine is not true, because it is the sentence on line 5, etc. None of these sentences can be successfully asserted, because none of them is true; but again I find myself slipping into nontruth: what I have *just* said is not true for it obviously includes the conjunction of these very same sentences; and also *this* last assertion is not true, and so on ad infinitum.

The innocent looking gap has developed into a black hole that sucks into itself every sentence asserting that some sentence in the hole is not true.[23] Formally, a hole hierarchy is definable as follows: S is a *0-hole* if it is a gap and, by induction: S is an $n + 1$-*hole* if it is an n-hole and 'S is true' and 'S is false' are gaps. Then S is a **black hole** if it is an n-hole for all n. (Here and elsewhere, quotes are to be read as semiquotes, whenever this is obvious from the text.) We shall use '*hole*', unqualified, to denote a 1-hole. Thus, we cannot assert of a hole that it is not true, or not false, or a gap (neither true nor false), without this second assertion being itself a gap.

---

[22] That some such minimal equivalence notion should obtain is inherent in the very assignment of truth values to sentence types. The giving up of (ii) would signify that something other than sentence type is involved in determining truth value.

[23] So far we assumed an equivalence satisfying the above mentioned conditions, and the following rules that imply that the line-1 sentence is a gap. (1) If 'S is true' is true then S is true. (2) If 'S is not true' is true then S is not true. (3) If the negation of a sentence is false then the sentence is true. We shall also assume, in the rest of the discussion, that if the negation of a sentence is true then the sentence is false and the equivalence of 'S is false' and 'not-S is true'.

Of course, syntactical properties can be asserted of *any* sentence. But no information about the truth value of a hole can be conveyed directly. And as for black holes, they are semantic untouchables. No semantic information about them can be conveyed in any way, be it as indirect and across as many layers as one may wish.

Actually, the arguments produced so far in the case of our example imply only that the sentence on line 1 is a "semiblack hole"; we have only iterated '. . . is not true' but not '. . . is not false'.[24] This is, however, bad enough. And it becomes a fully fledged black hole, under some additional very natural assumptions. In particular, it is easily seen that *every* gap is a black hole if we assume that *S* and '*S* is true' are equivalent, under an equivalence relation satisfying Leibniz's substitutivity law.[25] (Note that this is *not* the Tarski biconditional: '*S*' is true ↔ S, but only the requirement that '*S* is true' and *S* should get the same truth value, including *GAP*.) It is difficult to see how this equivalence can be avoided if truth values depend only on sentence types. Indeed, it holds in the system of Kripke (*op. cit.;* discovered independently around 1974 by Walther Kindt[26]), in its Martin-Woodruff[27] forerunner and in the variants of Gupta (*op. cit.*) and of Herzberger (*op. cit.*). Every truth-value gap is a black hole in these systems.

When truth values depend only on types, black holes are thus bound to occur on the most elementary levels of language. (Think how little is needed to produce sentences like the sentence on line 1.) Within such a system the little game I have just played, of repeatedly refuting myself by trying to assert the "unassertable," signifies a genuine metaphysical mystery. And yet the trick is quite transparent and we, Jack and Jill included, have no trouble at all asserting

[24] Our minimal assumptions (cf. *ibid.*) do not imply that 'The sentence on line 1 is false' is a gap. This can be shown by constructing a model—to be sure, a highly pathological one—satisfying the minimal assumptions, in which this last sentence is false

[25] That is to say, the equivalence of S_1 and S_2 entails that of 'S_1 is true' and 'S_2 is true' and similarly for 'false'. Actually, for the sentence on line n to be a black hole a weaker assumption than the equivalence of *S* and '*S* is true' will suffice. E.g., for $n = 2$ the assumption is that '*S* is true' is equivalent to '*S* is true' is true'. For larger n one iterates 'is true' accordingly, the assumption becomes weaker with each iteration

[26] "Introduction of the Truth Predicates into First-order Languages," in *Formal Semantics and Pragmatics for Natural Languages,* Guenthner and Schmidt, eds (Cambridge. Reidel, 1979).

[27] "On Representing 'True-in-L' in L," *Philosophia,* v (1975): 213–7.

this sort of "unassertable" by employing another token of the very same troublesome sentence. Language usage involves a mechanism for solving the black-hole dilemma. And if our goal is to model natural language, or a basic aspect of it, then the elimination of black holes—at least those of the simpler varieties—should be an overriding concern. Failure in this respect means that we have failed to capture something very essential to the functioning of language.

Although Kripke notices the failure, he plays it down, not giving the issue the central place it deserves.[28] "The sense [so he writes] in which we can say, in a natural language, that a liar sentence is not true must be thought of as associated with a later stage in the development of natural language, one in which the speakers reflect on the generation process leading to the minimal fixed point."

Indeed, the nontruth of the sentence on line 1 hinges on a sort of metareasoning. But this does not place the assertion on line 2 at a "later stage in the development of natural language." Certainly not at a stage that comes after the construction of the minimal fixed point, an inductive process that moves through all recursive ordinals! The fact that, after carrying out this transfinite construction, we are still left with the elementary problem of making sense of what Jack and Jill said shows that, far from having completed a home run, we are still at first base.

The basic (and simplest) version of my proposal eliminates the holes in all situations of finitary type—those in which every pointer calls only finitely many pointers.[29] To this type of network, which I call *locally finite*, belong all the tangled loops that can arise when finitely many people make assertions about each other, provided that altogether finitely many assertions are involved. In particular, it contains the examples discussed in the literature. The absence of holes is formally expressed by the following result, which holds for locally finite networks under quite general conditions:[30]

> If p has the value GAP, then there is a pointer q to the sentence $\neg Tr(p) \wedge \neg Fa(p)$ (i.e., the sentence asserting that p is a gap), such that q has the value T.

We can thus successfully assert, of any gap, that it is a gap. Another result, which applies under the same conditions, states that every sentence has a pointer to it that has a standard (T or F) value. (Recall that, by LPD, pointers to the same sentence, which get standard values, always agree in value.)

---

[28] This point has been nicely put by Burge (*op. cit.*).

[29] A pointer p calls a pointer q if one can reach q from p by one or more, finitely many, direct calls; cf. the "Preview" where the algorithm is described.

[30] The required condition is that there are infinitely many independent pointers to each sentence type, where "independent pointers" are those which have no common subpointers, a subpointer being the natural generalization of "subtoken."

In certain situations of infinitary type, when the language is sufficiently expressive, holes, even black holes, appear. Of these, some are eliminable by more sophisticated versions of the algorithm and it appears that further improvements are possible. It is far from clear how far one can go toward the elimination of holes. In the context of mathematical languages, the elimination of holes will bring us near to what one might consider a "universal language"—a self-contained system, one that includes its own semantics. For this reason alone it is highly plausible that rich languages must have holes. Note, however, that we have no formal proof of this. The standard argument from the liar is no longer valid for setups of the kind proposed here. But perhaps the question of proof should not even be raised at the present stage, since it is not clear what the present proposal might, in further developments, encompass.

In the broader perspectives of natural language "hole-like" phenomena are bound to occur. Some discourses take us to the edge of meaning, whereby we are tempted to express the inexpressible or to think the unthinkable. Holes are perhaps the inevitable price for a powerful language capable of evolving. Indeed, some situations can be described only by saying that we have changed the language, or that we have altered the very fabric of our conceptual framework. But I do not think that the Jack and Jill exchange calls for such a description. Here we can at least tidy up the more elementary levels of language.

A comparison with set theory comes naturally to mind. The analogy between the semantic paradoxes and those of set theory is well-known. Russell grouped them together and proposed the same kind of treatment.[31] His solution is based on a rigid predetermined classification of sets, which makes room for quantifying over sets of the same type, but not over sets in general. The hierarchy is incorporated into language itself. Axiomatic set theory, say ZF, has disposed of the built-in hierarchy, replacing the aggregate of concepts *set-of-type-i* by a single concept of set.

Russell's hierarchy of sets parallels his hierarchy of propositions, of which Tarski's hierarchy of metalanguages is a close relative. The representation of Tarski's hierarchy in a formal system yields an aggregate of truth predicates $true_i$, where i ranges over some list fixed in advance, but no single *true*.

---

[31] More recently, the analogy has been emphasized by Parsons Both types of paradoxes have been discussed by me as cases of what I called an infinity paradox: a contradiction that results from an attempt to construct self-containing systems. Cf. my "Paradoxes of Infinity and Self-application," *Erkenntnis*, XX (1983): 131–55.

In mathematical setups the problem is not acute. First, because 'true' is used there in a restricted way, which bars self-reference. Second, because, working inside set theory, we can assign truth values to mathematical statements—as long as these statements are about structures that are *sets*. A closer look at the second point shows that the metalinguistic hierarchy has been subsumed under the hierarchy of sets (the one obtained by iterating the power-set operation). Problems resurge when the structures in question are no longer sets; in particular, when truth in "Cantor's universe" is considered. In this respect, axiomatic set theory has not solved the problems arising out of the set-theoretic paradoxes. It has only shifted them to the theory's fringes, where they are of no concern, except to set theoreticians who investigate the theory's borders (for a detailed discussion, cf. *ibid.*).

This perspective marks a clear goal for the semantics of 'true'. A semantics of formal languages—containing their own truth predicate—should do for the concept of truth what set theory, such as ZF, does for the concept of set. By minimizing the black-hole phenomena we are shifting the problematic points, where the system's inadequacies show, to the less accessible parts of the system.

Whereas set theory is informed by mathematical thought, pointer semantics draws directly on insights derived from natural language. (Mathematics itself, however, is an outgrowth of thought manifested in ordinary language.) Yet it results in systems that are as formally precise as any piece of mathematics. The present work can be seen as a first step that eliminates black holes at the basic linguistic level. As mentioned earlier, black holes reappear, in highly expressive languages. These are cases in which the pointing relation itself is describable in the language and Gödel's techniques are available. Attempts to eliminate such black holes lead to more sophisticated versions of the algorithm and to some intricate, interesting questions. Thus, the formalization of mechanisms inherent in linguistic practice also gives rise to foundational questions of logic.

III. TOKENS, TYPES, AND COMPOSITIONAL SEMANTICS

Let me start with a commonplace observation. When a child learns a language he encounters first of all tokens of linguistic entities. The child acquires the language by learning the effects of token display and by participating in the token-display exchange. This observation does not prove that tokens are prior (in some nontrivial sense) to types; for the ability to learn the language is preconditioned by the ability to recognize token similarity and, thereby, to classify tokens according to type. The observation is made in order to underline the obvious fact that there are many ways of passing information by means of an interactive game-like process. A useful description of the game must, inevitably, ignore through abstraction various aspects; otherwise, we shall be submerged under sheer detail.[32] Of course, we may ignore only inasmuch as it does not restrict us seriously in accounting for what is going on. There are thus two conflicting requirements: that the account we give be sufficiently simple, and that it be adequate for handling a broad enough class of applications. Usually there is trade-off between the two.

The fixing on sentence types, as on the basic truth-carrier units, is a natural move. It leads to an enormous simplification, and it clears the way for compositional semantics: a semantics whereby linguistic entities are analyzed as complexes whose meanings derive in a systematic way from that of their components, via certain—relatively few—recursive rules. The truth value of a sentence is thus determined recursively; the context in which the sentence occurs can be completely ignored.

A description of this kind fits accurately formal languages constructed by logicians; which is no wonder, since it is the very principle on which such languages are constructed. It does not fit natural languages, where context dependencies abound, due to indexicals, demonstratives, and ambiguous terms. Nonetheless, it reflects a very essential aspect of natural language. For if the dependency of truth value on context can be traced to the shifting denotations of certain context-sensitive terms, then once these denotations have been determined—how, is another matter—we can derive truth values by means of the above mentioned recursive rules.[33]

Call two tokens *dissenting* if they are tokens of the same sentence type, but are accorded different truth values. Then the account just outlined amounts to the following. If p and q are dissenting tokens, then the sentence type should contain terms, say $t_1, t_2, \ldots, t_n$, whose references differ on the two occasions. If we are to replace each t_i by a suitable t_i', which denotes on both occasions what t_i denotes in the context of p, the resulting tokens, say p' and q', would have the same truth value, namely, that of p. We can call such an elimination of context dependency *local disambiguation*.[34]

---

[32] On a more fundamental level, the very concept of description means that many details have been ignored

[33] Such a view underlies Davidson's semantic program

For example, when 'I am tall' is uttered by two speakers, only one of whom is tall, the conflict in truth value is traceable to the different denotations of 'I'. The dissent can be resolved by using proper names. If needed, a name can be introduced by stipulation: "Let S1 be the first speaker." In the counterfactual situation, where the two speakers use 'S1' instead of 'I', the two utterances have the same truth value: that of the first speaker's original utterance.[35] Or consider two tokens of 'tanks are never made of wood', one in a treatise on weapons, the other in an essay on storing methods. The difference in truth value is due to 'tanks', and is eliminable by substituting the term by 'armored vehicles on caterpillar treads'.[36]

The picture is radically different when every term contained in the sentence type has the same denotation, on both occurrences in two dissenting tokens. Here, a truth-value difference cannot be reduced to denotational shifts of sentence parts. It is accounted only by the token's role as a unit within some global arrangement. And such, I hope to show, are the cases deriving from the self-referential usage of 'true'. The argument for the failure of indexicality shows, in fact, the failure of the usual compositional semantics.[37]

First, I take it that in the two-line puzzle, and in similar cases, we *do* have token dissent. We do want to say that the sentence on line 1 is not true but the one on line 2 is; we do side with Jill against Jack.

---

[34] The "context insensitivity" of the terms t_i' is not necessarily absolute. It only means that they have the same denotations throughout some wider context, which includes that of p and that of q.

The terms t_i can be either noun phrases or predicates. I assume here, for the sake of simplicity, extensional contexts. But this is not essential. Local disambiguation makes sense in intensional settings, provided that we apply some stronger equivalence notion. The term t_i' should be intensionally equivalent to t_i-in-the-context-of-p.

[35] Instead of changing, counterfactually, the original utterances, we can let the two speakers reutter the sentence, with 'I' replaced by 'S1'; the natural assumption being that, in the given circumstance, an utterance and its repetition by the same speaker are equivalent. This strategy does not work if the indexical is time-sensitive, unless the repetition is sufficiently close so as to make the time difference insignificant.

[36] In lumping together indexicals with cases of ambiguity, I am not glossing over what distinguishes the first. The "ambiguity" of 'I' means that its denotation shifts according to context. Because the shift is systematic, 'I' can be regarded, in one sense of 'meaning', as having a fixed meaning: it always denotes the speaker. Where no such systematic account is available, we must rest content with what informal pragmatics will provide.

[37] The qualifier 'usual' is added here, in order not to rule out other types of compositionality. For example, one may consider composition of networks.

Or, to use Buridan's example,[38] if Plato says at time *t*: "What Aristotle says at time *t* is not true," and Aristotle says, at the same time, the same about Plato, then both sayings are not true; but if a third party were to assert at that time that what Plato (or what Aristotle) says is not true, *his* saying would be true.

Evidently, 'The sentence on line 1' in the two-line puzzle refers in both tokens to the same object, be this sentence type or sentence token. The only candidate whose meaning, or denotation, might have changed is 'true'. But note that the argument for the nontruth of the line-1 sentence employs 'true' in the very same sense of the 'true' occurring on line 1.[39] The argument simply shows that the sentence on line 1 does not fall under what 'true' *in this very sentence token* denotes. And if *this* is what we write on line 2, then the 'true' on the second line has the same denotation as that on the first. (Indeed, if the 'true' on line 2 is construed as being on a higher level, then the sentence on line 2 is false!) The difference in truth value is not caused by some shift involving the 'true' that occurs in the sentence, because no shift takes place.

Interestingly, the analysis just offered accords with that given by Burge in his discussion of a similar puzzle (*op. cit.*). Phrased in the present terms, his conclusion is that the tokens of 'true' on the two lines should be construed as standing for the same truth concept. The difference in truth values is traced by him to different implicatures involved in evaluating the sentence. (As will be shown in section V, some unspecified implicature mechanism is implicitly presupposed by Burge throughout, while his indexicality account does no actual work.)

Although I find the foregoing argument sufficient for my purpose, I shall supplement it by another, which is, I think, of independent interest. Again, I consider the two-line puzzle. I shall not appeal to how we derive the conclusion about the line-1 sentence, however, but proceed in a more general way. Let the 'true' on line 1 be changed to a nonindexical 'true*' that denotes once and for all what 'true' in the original context denotes.[40] We thus freeze 'true' in its line-1 context. The freezing, one feels, should not affect anything that has to do with truth. Saying, while addressing Jack, 'you are tall' and saying 'Jack is tall' come to the same.

---

[38] Quoted in Burge, op. cit.

[39] The argument rests on (T1). If '*x* is not true' is true, than *x* is not true. (And in this phrasing the various occurrences of 'true' denote the same truth concept, whatever that may be.) The familiar argument now runs thus. If the line-1 sentence falls under what 'true' in the line-1 context denotes, then by (T1) that sentence does not fall under what 'true' in the line-1 context denotes.

[40] Either the change is effected by erasing and rewriting, or we imagine a counterfactual situation in which the second version is written instead of the first.

There is yet a subtle complication: By changing 'true' to 'true*' we have altered the reference of 'the sentence on line 1'. There is thus the additional change in what that noun phrase denotes. Nonetheless, as far as truth or nontruth is concerned, this added difference should not matter. It might have, had the sentence said something about lexical properties. Changing

> You are tall and this last utterance of mine is short.

said by Jill to Jack, to

> Jack Horatio MacWiltington the Third is tall and this last utterance of mine is short.

makes a difference. But with 'true' instead of 'short', the two come to the same.[41]

Since the original line-1 sentence is not true, its frozen version is not true as well. And since we can assume that truth* satisfies (T1) of the "Preview," this being a minimal requirement of any truth concept, it obviously follows (cf. fn. 39) that the frozen version is not true*. Having written this last conclusion on line 2, we get another token of the frozen version, this time a true one. We have eliminated the indexical context dependency, but we get back two dissenting tokens. Local disambiguation fails.

Thus, indexicality, or anything like it, cannot account for truth-value changes accruing through the self-referential usage of 'true'. This is not to say that the picture of truth as involving meaning variations is altogether wrong. Something like truth-level shift does take place. And it is important to disentangle this true aspect from the wrong idea that 'true' is indexical.

We can, in fact, view the judging of the sentence on line 2 as taking place on a higher level than the judging of the sentence on line 1. But the shift to higher level is determined directly by the place of the token in the network, not by the mediation of context-sensitive terms in the sentence, because all the terms retain their original meanings. We have seen that the evaluation procedure provides a formal reconstruction of the idea of levels. The line-2 token is seen to be one level higher, because its evaluation takes one application of a jump rule, while no jump rule is needed for evaluating the first. If one wishes, one can get, as a corollary of the level assignment to pointers, a level assignment to 'true'. The prescription is simple. Assign to the occurrence of 'true' in a token of 'true(p)' the ordinal number assigned to p; for this is the level that the evaluation of p reaches. In particular, the 'true' on both lines is assigned level 0—the number assigned to the line-1 token. But were we to write on a third line 'the sentence on line 2 is true' (or 'not true'), *that* 'true' would get level 1.

---

[41] The analogy is not altogether complete. With the changes in two separate conjuncts: 'you are tall' and 'this last utterance of mine is true', we can argue formally by considering first one, then the other. Such a procedure is not at hand when the changes are fused in the same atomic sentence. But this cannot provide any ground for taking the change of 'true' to 'true*' as a cause for truth-value change. Whoever has misgivings on this score may consider a variant of the liar in which the sentence is, so to speak, split: on line 0 we have, 'The sentence on line 1 is true', and on line 1, 'The sentence on line 0 is not true'. Can the freezing of 'true' on line 1 make a difference? I cannot see that it should matter for the sentence on line 0 and, by consequence, for the sentence on line 1 either And neither should an additional freezing of 'true' on line 0

Occasionally, truth levels are indicated by attaching subscripts to 'true'. This would be harmless, provided that we are not misled to see in these subscripts, or indices (as they are also called), some form of indexicality. The indices play no role in determining truth value. They come after the fact, as an annotation to a finished process, an underlining of certain aspects of the evaluation.

While the reference of 'I' is picked through the indexical mechanism from the multitude of humans, nothing of the kind obtains in the case of 'true'. There is no preexisting multitude, or hierarchy, of truth concepts, out of which we pick the concept we invoke. The hierarchy is created by the evaluation process itself, by the very act of judging, by the linguistic exchange.

The functioning of 'true' exemplifies that aspect of linguistic activity whereby sentence tokens function as the units of the exchange. The meaning of a token is not reducible to what the sentence says, even when the denotations of its terms have been fixed. It derives also from the token's particular place in a global setup.

The principle, under other forms and on different scales, is quite common. An event, an assertion, a work of art might mean whatever it means by virtue of its place in a global system of similar events, acts, or works. This is particularly true in the artistic domain. Duchamp's famous urinal refers to the tradition of art exhibitions. The reference, though not formal and not explicit, is there. For the place of that particular display within the historic network of art is what endows Duchamp's exhibit with artistic meaning. Moreover, it would be erroneous to derive this meaning from the urinal itself, as if it were an "indexical" that indicates different things in different contexts; quite the contrary. That object means or connotes, within the displaying event, whatever a urinal means or connotes in everyday life. Only by retaining its ordinary, undignified status, can it cause the surprise, the outrage, and the joke that constitute the aesthetics of the display.

We can now take stock and assess the extent to which pointer semantics departs from a type-based one. We can use as a sort of measure the extent of token—or pointer—dissent caused by the functioning of 'true' (and not by indexicals, or ambiguities of sorts). In this respect, the formal system of the present proposal provides a good view, as it excludes indexicals, or any kind of ambiguity. Now, the results LPD and HLD mentioned in the "Preview" imply a very simple picture, perhaps surprisingly simple—given that it holds for full first-order logic and arbitrary pointer networks. Whenever pointers dissent, one of them has the value *GAP* and the other, which has a standard value, is one level higher in the linguistic hierarchy. Informally, we can say that departure from type-based semantics occurs just when a pointer fails to express that which another at a higher level succeeds in expressing.[42] Thus, departure takes place when a necessary ascent in the linguistic hierarchy cannot be accommodated within Tarskian semantics or any of its offshoots—and here I include Kripke's system and all its variants—thereby forcing us to change the language. To encompass it within a single language, which is what the present framework does, requires a price and the results show that we are paying the minimum.[43]

While the semantic paradoxes have focused a great amount of research, resulting in an extended and often sophisticated literature, they have not been at the center of the research program of natural-language semantics. This is understandable, given the tangle of problems encountered in the rather elementary stages of the investigation. To give the semantics of 'true' while we are far from clear on that of 'event', or that of mass terms, might seem a luxury.[44] But the tendency to regard the problems raised by the paradoxes as something esoteric, to be handled later (if at all) when more basic aspects have been clarified, is, I think, an error. We are not concerned here with marginal brain teasers, but with an essential feature of language. The failure of compositional semantics (where "composition" is intended for sentence parts) is highly significant.

---

[42] In this respect, all such cases are like the simple two-line example. But the sentence in question can be as complex as any first-order sentence and the failure can be caused by an extremely intricate loop. Moreover, it can occur at any level, the ordinal of which can be of any cardinality not exceeding that of the total network.

[43] I should add, however, that it is conceivable that in more sophisticated variants of the algorithm LPD will fail I see no reason why LPD, pleasing as it is, should be imposed as a prerequisite on evaluation procedures in general. With indexicals present, different tokens of the same sentence may get different non-*GAP* truth values. That a similar phenomenon may arise because of deeper reasons, deriving from the use of 'true', is not to be ruled out.

[44] *Using* the concept of truth in order to give semantics in the form of truth conditions, as was proposed by Davidson, is of course quite different from giving the semantics of 'true' itself.

Note also that 'true' is *the* paradigm of a network-creating predicate. It can send us, via 'what she said', 'whatever he asserted', 'what is written on page 1', and their like, from John to Joan, from Joan to Jack, to Jill, to McX, and so on; the tour is unrestricted and sometimes endless. 'Know' (in the sense of "know that") and 'believe' (in the sense of "believe that") have similar potential: 'I know what she knows', 'what he believes I doubt'. But 'true' is the more fundamental.[45] It is no wonder that problems arising out of this chain-forming capacity have surfaced in the theoretical research of artificial intelligence. Once it is realized that knowledge, or belief, are to be modeled as predicates rather than as sentential operants,[46] the semantics of 'true' becomes relevant in the contexts of knowledge and belief representation. This can be seen in quite a few works published mostly in the volumes of TARK. Thus, L. Morgernstern[47] and M. Kremmer[48] draw on Kripke's model, while N. Asher and H. Kamp[49] follow the constructions of Gupta and Herzberger. Among other works motivated by problems of self-reference in epistemic contexts are those of Asher,[50] R. Koons,[51] D. Perlis,[52] J. des Rivieres and H. Levesque,[53] and R. Thomason.[54]

---

[45] Note how often 'know', and to lesser extent 'believe', ride on the back of 'true'. We might say 'I know that what she says is true', but we cannot say 'I know that what she says', while 'I know what she says' means something different. Knowledge modalities like $K_J(x)$, which are used in formal systems, are rendered into natural language by invoking 'true' or an equivalent expression: '*J* knows that *x* is true' or '*J* knows that *x* is the case'.

[46] Sentential operants are too restrictive. 'John knows something that Bill does not know' is inexpressible if 'know' is treated as a sentential operant.

[47] "A First Order Theory of Planning, Knowledge, and Action," in TARK, J Halpern, ed. (Los Angeles: Morgan Kauffman, 1986), pp. 99–115

[48] *Logic and Truth,* Ph.D. Dissertation (University of Pittsburgh, 1986).

[49] "The Knowers Paradox and Representational Theories of Attitudes," in TARK (1986), pp. 131–48

[50] "Reasoning about Belief and Knowledge with Self-reference and Time," in TARK (1988), pp. 61–81. Also, "Intentional Paradoxes and an Inductive Theory of Propositional Quantification," in TARK, R. Parikh, ed. (Los Angeles: Morgan Kauffmann, 1990), pp 11–28

[51] "Doxastic Paradoxes without Self-reference," in TARK, M. Vardi, ed. (Los Angeles: Morgan Kauffmann, 1979), pp. 29–42.

[52] "Languages with Self-reference I: Foundations," *Artificial Intelligence,* xxv (1985) 301–22.

[53] "The Consistency of Syntactical Treatment of Knowledge," in TARK (1986), pp. 115–130.

[54] "Paradoxes and Semantic Representation," in TARK (1986), pp. 225–39.

IV. APROPOS PROPOSITIONS

Some would have it that truth and falsity attach to propositions (or statements)—entities from which linguistic particulars have been abstracted away. There is, however, no agreed view on the nature of these objects and the difficulties surrounding the concept may incline us to consider, at least provisionally, less problematic creatures. In any case and however propositions are conceived, the crucial question is that of relating them to linguistic constructs—to sentences, for example. On the present view, the question takes the form: What proposition, or statement,[55] is expressed by a pointer?

I have suggested in the "Preview" that, prima facie, a pointer can be taken to express whatever is expressed by the sentence it points to—except for failures, marked by *GAP*. On this suggestion, non-failed pointers to the same sentence express the same thing. Since (by LPD) such pointers get, invariably, the same truth value, we will not run into trouble, as far as truth values are concerned. The plausibility of the suggestion rests, however, not on LPD, but rather on its proof, which shows the following. Whenever p and q are non-failed pointers to the same sentence, their evaluation requires the same sequence of necessary steps, ending at a stage where each can be assigned the same truth value by applying the same rule.[56] This indicates that an account that correlates propositions with sentences could be extended to pointers. The heuristics would be as follows. Two non-*GAP* pointers express the same proposition, inasmuch as this can be said of the sentences they point to.

---

[55] There is considerable overlap in the use of the two terms and I am not altogether certain as to the "right" one. There is a usage, which I do not follow, whereby propositions are taken as the meanings of declarative sentence types, including indexical ones. By contrast, statements are attached to particular utterances. In the tradition deriving from Frege and Russell, however, the proposition expressed, say, by 'it is now raining' changes according to time. Read thus, propositions can be associated with sentence types only after converting the sentences into "eternal" versions. What is then the difference between this notion of proposition and the notion of statement as advocated by J. L. Austin and Peter Strawson? It appears that propositions are prelinguistic entities, existing in their own right, whereas a statement is something one makes by producing a linguistic token. Put bluntly, if there were no people, there would have been no statements but there would have been propositions. Stripping them of their more metaphysical aspects, one can say that propositions are possible statements. And this is how I propose to view them here. I have opted for 'proposition', because it is usually this term which is employed when a formal treatment is intended. Thus, Barwise and Etchemendy (*op. cit.*) propose a formal model of so-called Austinian propositions for what Austin, who rejected 'proposition', called statements. Indeed, if propositions are possible statements, then the general theory of statements is the theory of propositions.

[56] Were we to use another version of the evaluation procedure, a similar analysis of the procedure would be required in order to justify stipulations about "expressing the same thing."

The propositions to be gotten along the suggested lines are "ontological" rather than "epistemic." A person stating what to him is a perfectly clear thought might fail to express a proposition, because unwittingly he has involved himself in some loop. This is similar to the view obtained from some current accounts of propositions. On these accounts, certain terms (the so-called directly referring ones) introduce their very references as components of the proposition.[57] Mistaken beliefs about reference may lead to a person's expressing a proposition he had not intended. And when a directly referring term fails to refer, no proposition is expressed.

But nothing prevents us, at least not in principle, from considering other kinds of propositions aimed at reflecting finer shades of meaning. Just as, for various purposes, it is natural to hide the reference-picking machinery of certain terms, it might be desirable, for other purposes, to expose it in the proposition. What propositions are depends on what dimensions of meaning they are supposed to reflect. The question what are the right propositions is as meaningful as the question what sort of items a map should display, or what is the right granularity of a film.

A move to more epistemic propositions may entail distinctions between various gap pointers. The gap constituted by the line-1 sentence derives from a contingent fact: the place where that sentence is written. Someone might have put it there because of a mistaken belief about line numbers.[58] By contrast, the loop produced by 'What I am saying at this very moment is not true' cannot be amended by moving it, and other utterances, around. Assuming that the speaker knows the language, it is difficult to see what statement of fact he might have intended. The present formalism does not distinguish between the two; both are construed as a pointer, p, which points to $\neg Tr(p)$. The distinction can be introduced at the cost of additional structure: for example, a set of possible pointing functions in addition to the actual one.[59]

Finally, it can be argued that *some* meaning is expressed even in the case of noncontingent gaps. For a pointer provides a prescription for computing truth values. It amounts to a kind of algorithm. And as an algorithm it makes sense, even when it leads inevitably to a nonterminating computation.

---

[57] Cf. Kaplan, "Demonstratives. An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in *Themes From Kaplan*, J. Almog, J Perry, and H. Wettstein, eds. (New York Oxford, 1989).

[58] In an example of Prior, used by Burge, a student who thinks he is in room 9 writes on the blackboard: 'Whatever is written on the blackboard in room 10 is false', expressing thereby his true belief. It turns out that he himself is in room 10. (N. Prior, "On a Family of Paradoxes," *Notre Dame Journal of Formal Logic*, II, 1 (1961): 16–32.)

[59] Distinctions of similar nature, between kinds of pointing, are used in certain more sophisticated variants of the evaluation procedure, aimed at reducing black holes

While not endorsing, at least initially, any particular position concerning propositions, the present proposal incorporates a certain methodology: consider propositions after having clarified sufficiently the rules and mechanisms of truth-value assignments. Propositions—if and when we introduce them—can be, then, tailored to fit the patterns we deem important. To do otherwise is to commit ourselves to preconceived notions, which can be interesting and fruitful in their domain of origin, but ill-matched to the issues at hand.

An opposite approach underlies the work of Barwise and Etchemendy (cf. fn. 15), some aspects of which I shall now discuss in brief. Barwise and Etchemendy propose formal models of propositions, constructed according to situation semantics. They classify their propositions into Russellian and so-called Austinian ones, with clear preference for the latter. One of their goals is to include in the model *cyclic* propositions: those involving self-reference. And their ingenious idea is to employ for that purpose AFA (set theory with Aczell's *antifoundation axiom*)—a theory of non-well-founded sets which accommodates, naturally, a variety of "cyclic creatures," for example, a proposition P such that $P = $ *not-P*.

While the BE (Barwise-Etchemendy) framework assigns truth values to propositions, it does not provide guidelines, let alone a procedure, for assigning truth values to sentences—either types or tokens. For when it comes to determining the proposition that is expressed by a given utterance, a kind of free parameter enters.[60] In cases of self-reference there is considerable latitude; much is up to the interpreter, who can toggle truth values by suitable adjustments in what he can choose. Therefore, such utterances are construed as ambiguous. Presumably, Barwise and Etchemendy regard the problem of disambiguation (which must be addressed if we are to determine truth values) as belonging to some branch of pragmatics.

It can be shown that the assignment of truth values according to the principles of pointer semantics is outside the scope of the BE framework. For the needed distinctions cannot be made in AFA.

---

[60] Namely, the situation—in the technical sense of situation semantics—associated with the utterance. A situation can be very roughly described as a collection of facts. But the complete circumstances of the utterance—including, if you wish, the whole world history—fall short of constituting a complete situation For the truth or falsity of some circular proposition is itself a "fact" that the interpreter may, or may not, include in the situation.

For example, AFA cannot provide for distinguishing between the sentence tokens on lines 1 and 2 in the two-line puzzle.[61]

Can we correlate sentence tokens with propositions of the BE kind, whose truth values are the values derived by pointer semantics? To a certain extent this is possible. It may work in general, but it requires the removal of certain restrictions underlying the BE modeling of propositions.[62] Perhaps these are artifacts of the modeling.

The major nontechnical point of conflict between the BE framework and pointer semantics is the former's failure to make place for the notion of a *failed utterance*. It provides the liar sentence and its kin with circular propositions, which I do not find objectionable—as long as we mark certain propositions by *GAP*. But, in the BE framework, propositions are either true or false. When $P = not\text{-}P$, the proposition and its negation are accorded *F*; which fact is explained in BE by the special modeling of negation—clearly not the classical concept. Under a different negation concept, proposed in BE and called *denial*, both a proposition and its denial can be true. Whatever the insights revealed by these nonstandard negation concepts, their service in disguising gaps is negative. A picture that fails to distinguish certain types of failure can be quite misleading. It is also at odds with Austin, who devoted much effort to the taxonomy of failures and who distinguished failures caused by illegitimate self-reference from mere self-contradiction.[63]

V. THREE PREVIOUS ACCOUNTS

I shall discuss three accounts related to my proposal, those of Parsons (which can be seen as a development of a theme in Russell),

---

[61] The two nodes in figure 2 of the "Preview" represent in AFA the same set, namely, the unique set containing 1 as its sole member This is forced by the extensionality axiom

[62] The definition of disjunctive propositions is such that all the disjuncts are about the same situation—which is also the situation the disjunction is about. A similar restriction applies to conjunctions. This has some problematic consequences. Consider a disjunction $S_1 \vee S_2$, where S_1 is a liar sentence and S_2 is a sentence saying that S_1 is not true. If both disjuncts must be about the same situation, the second disjunct, and along with it the whole disjunction, come out nontrue. In pointer semantics this disjunction (i.e., a token, or some pointer to it) gets the value *T*, because the second disjunct does

[63] "Those of the so-called 'logical paradoxes' . which concern 'true' and 'false' are *not* to be reduced to cases of self-contradiction, any more than '*S* but I do not believe it' is. A statement to the effect that it is itself true is every bit as absurd as the one to the effect that it is itself false There are other types of sentence that offend against the fundamental conditions of all communication in ways distinct from the way in which 'This is red and this is not red' offends . . ." Austin, "Truth," in *Philosophical Papers* (New York: Oxford, 1961). Fn. in the 1970 paperback ed , p 129.

Burge, and Skyrms.[64] These accounts do not assign truth values to sentence tokens, but they share with the present work the view that places the strong liar and its kin at the heart of the semantic paradoxes. Needless to say, I have no intention to review the vast literature devoted to the liar. The choice of these works is dictated by their relevance to my proposal and reflects no attempt at ranking.

Ambiguity and interpretation schemes. Parsons proposes to view the liar and its kin as cases in which different interpretation schemes, based on different truth concepts, are applied to the same sentence (type). No formal explication is given to "interpretation scheme," but a direction is suggested by rewriting '*A* is true' as 'there exists a proposition *P* such that *A* expresses *P* and *P* is true'. The interpretation scheme is then determined by fixing the domain of propositions over which the quantified '*P*' ranges. The idea goes back to Russell, who suggested that the paradox arises from an attempt to quantify over "too many propositions" and is to be resolved by restricting the quantifier ranges. Such a solution underlies Russell's theory of types. Since there are many natural domains that can serve as the range for the propositional variable, the sentence turns out to be ambiguous. But whatever the domain, it cannot include the proposition expressed by the liar sentence itself.

Intuitively, the interpretation scheme reflects the comprehensiveness of the point of view from which truth-value judgments are made. In the case of our two-line puzzle, the sentence on line 1 fails to express a proposition, under the first "straightforward" interpretation; but under a second scheme, which takes into account the previous failure, the sentence is seen to be true. Parsons does not give an account of how the interpretation is determined, except for noting that it is determined by the circumstances of the sentence's utterance. He proposes to regard this ambiguity in the same light as the ambiguity surrounding the interpretation of quantifiers or the use of 'say', 'mean', and their like.

In a 1982 postscript to his paper, Parsons suggests that Burge's indexical conception of 'true' can serve to explicate the ambiguity.

---

[64] Parsons, "The Liar Paradox," *Journal of Philosophical Logic*, III (1974): 381–412

Russell, "Mathematical Logic as Based on the Theory of Types," in *Logic and Knowledge: Essays 1901–1950*, R. Marsh, ed. (New York: Macmillan, 1956), pp. 59–102.

Burge, *op. cit.*

Skyrms, "Return of the Liar, Three-Valued Logic and the Nature of Truth," *American Philosophical Quarterly*, VII (1970). 153–61; and his "Intensional Aspects of Semantical Self-reference," in *Recent Essays on Truth and the Liar Paradox*.

But, in the light of the analysis of section III (and of the next subsection devoted to Burge's work), this would not do justice to the intended picture. For the ambiguity of the liar is not to be traced to indexical-like shifts in what the 'true' that occurs in the sentence denotes. If there is ambiguity, it can derive only from different points of view that we—the interpreters—adopt from the outside.

According to pointer semantics, there is no ambiguity. The token on line 1 is definitely not true (and not false either); the token on line 2 definitely is. But if one were to smudge over the distinction between tokens, one could say that the sentence *type* is ambiguous: different instantiations of it get different truth values. One can therefore regard different pointers to the same sentence as expressing its meaning under different interpretation schemes.

There are other systems that can provide formal backups for the ambiguity picture. The constructions of Gupta and Herzberger, whereby the liar changes its truth value according to the evaluation stage, give rise to such a picture. And the more recent work of Barwise and Etchemendy, which I discussed above, provides another modeling of ambiguity. It is indeed directly related to Russell's ideas, since one of its aims is to model Russellian propositions.

While these models do not solve the two-line puzzle, they solve or illuminate what we might call the one-line puzzle, which is as follows. A philosopher contemplates the sentence on line 1, which asserts its own nontruth. First he infers that the sentence is not true. Realizing on second reflection that the sentence itself "says" just this, he infers that it is true after all. He may continue to the third round, whereby the sentence becomes nontrue, after which it becomes true again, and so on.

Burge's proposal. I have already discussed in section II the indexical construing of 'true'. Here I shall confine myself to more specific details of Burge's work. His proposed setup comes in two parts. The first consists in a formal language containing a family of truth predicates, true_i, $i = 0, 1, \ldots$, arranged in a Tarski-like hierarchy, with subscripts denoting the levels. Each predicate is to serve as a truth predicate for the portion of the language "below it," that is, for all sentence types containing truth predicates with a strictly smaller index. Yet each predicate takes as arguments all sentence names of the language. Because of self-reference there are, for each i, sentences that inevitably fail to get a truth_i value: neither the sentence nor its negation is true_i. The idea is to regulate the assignment of truth_i values so as to approximate, as far as possible, the properties required of a truth predicate, while maintaining the predicate's position in the hierarchy. Burge stipulates various plausible axioms; for example, axioms to the effect that if A is true$_i$, then A, as well as *true*$_i$ (A) are true$_k$, for all $k > i$.[65] Apparently, the proposed axioms are not sufficient.[66] But this is a technicality that can be amended. There is, in fact, a neat generalization of Kripke's construction to languages containing a hierarchy of truth predicates, which achieves the desired result in a simple straightforward way. It is mentioned briefly in my 1983 paper.[67] We can therefore assume that truth values have been assigned in a satisfactory way to all sentences and all i. Call this language the stratified language. (Note that a liar$_i$, i.e., a sentence saying of itself that it is not true$_i$, is not true$_i$ and neither is its negation. But a liar$_i$ is true$_i$ when evaluated from the outside, via the usual truth definition. Since each *true*$_k$ behaves like an ordinary truth predicate for the portion of the language below it, a liar$_i$ is true$_k$ for all $k > i$.)

The second part of Burge's proposal is to rewrite the ordinary sentences within a given discourse as sentences of the stratified language; this is done by "indexing," i.e., by attaching subscripts to the occurrences of 'true'. No precise rules are given; Burge sees it as a pragmatic question and provides certain informal guidelines.[68] The recourse to informal pragmatics has brought the proposal under severe criticism by Gupta.

The main trouble with Burge's proposal, however, is not that the pragmatic principles for indexing are imprecise, but that the indices do not do the desired work. An indexing can only lead to ***truth**$_i$* values; it does not determine at which level a given sentence is to be evaluated: Which *true*$_i$ are *we* employing when we say of a sentence *A*—written already in indexed form—that it is true, or not true, or false? Burge invokes here what he calls a pragmatic implicature: consider the evaluating sentence—the sentence asserting the truth, or nontruth, of *A*; index 'true' in *that* sentence. That index is the level at which *A* is evaluated. This, however, is only the beginning of the story; for occasionally the implicature is canceled in favor of a different one. For example, the sentence on line 1 (in the two-line puzzle[69]), which asserts its own nontruth, is the evaluating sentence for itself. Suppose the 'true' there is indexed by *i*. By the implicature, when asserting that the sentence is not true, we are—in fact—asserting that it is not true$_i$; which, indeed, it is not. But the same implicature rules that when we evaluate our assertion, we do so again on level *i*. Therefore, our assertion, expressed by the sentence on line 2, is nontrue as well. This unwelcome result is avoided by having the implicature cancelled and by raising the level of the second evaluation to some *k*, greater than *i*. As Burge sees it, the move from line 1 to line 2 signifies a switch from one implicature to another. The change of truth value is only apparent: both are true$_k$, and both are not true$_i$.

---

[65] Actually, Burge considers various sets of axioms, but this need not concern us here.

[66] The following is, I think, not derivable: If $\neg$*true*$_i$(A) is true, then A is not true$_i$.

[67] I found it when, independently of Burge's work, I was considering formal languages that incorporate Tarskian hierarchies. Here it is. Start by evaluating all sentences by the *standard rules*, treating each of the predicates *true*$_i$ as a truth predicate. The standard rules consist of the usual (Tarskian) truth conditions (applied, as in Kripke's construction, via Kleene's strong three-valued logic) as well as the following rule If A is true$_i$, then *true*$_i$(A) is true$_j$ for all j (including the case $j < i$!). Having done this, extend the resulting valuation as follows. For each A which is not true$_0$, declare the sentence $\neg$*true*$_0$(A) to be true$_i$, for all $i > 0$ Now repeat the same process of applying the standard rules, but only to the truth predicates on level > 0. Thus, by the end of the first stage the extension of *true*$_0$ is determined. At the end of the second stage, for each A which is not true$_1$, declare $\neg$*true*$_1$(A) to be true$_i$ for all $i > 1$. Again, repeat the process for all predicates at level > 1. And so on, until we run out of truth predicates

[68] A principle called "verity" instructs us to assign indices in a way that maximizes the applicability of the truth-defining schema, that is, in a way that minimizes gaps. A less important principle called "justice" says that the assignment should not prefer one statement to another, unless for some good reason. Thus, if each of two persons asserts of the other that he is not saying the truth, both should be given—for reasons of symmetry—the same subscript, resulting in both being gaps Either principle is subject to a *ceteris paribus* clause

What Burge fails to observe is that the resort to implicatures voids the indexing of its significance. Not only the liar, but all problematic cases of self-reference, require implicature switches. Thus, in the Plato-Aristotle example quoted from Buridan (cf. section III), the 'true' in both sayings gets the same index, resulting in both being nontrue. If a third party were to assert that Plato's, or Aristotle's, saying is not true, *his* saying would be true. But this can be sustained only if the implicature employed in judging the first two assertions is changed when judging the third. Or take Nixon-Dean example, where each says of the other that all his utterances concerning Watergate are not true. If all other Watergate assertions made by each are not true, a loop is created and *we* can say that these two assertions are not true. Again, the implicature must be changed when judging *our* assertion; because when Nixon's and Dean's sayings are spelled out as conjunctions, our assertion is seen to repeat a conjunct from each.

In fact, whenever the jump rule is used in the evaluation procedure, an implicature must be canceled in favor of a different one that shifts the level of 'true' in the evaluating sentence. Since the shifts are cumulative, a count must be kept to determine the "right implicature" for choosing the level; and the level can be any finite or infinite ordinal. Burge ignores all this. Having introduced implicature in the analysis of the liar, he goes on in the second half of the paper to discuss the stratified language and the indexing, without mentioning that matter again. The process of implicature shifts is altogether obscure; no guideline, nor any indication, is given. But there lies the real problem. For knowing the mechanism that governs implicature choice is almost like knowing the evaluation procedure. It is this mechanism which does the real work. Indexicality rests idle.

---

[69] The example used by Burge is somewhat different, but the difference does not matter here.

The intensional conception. The analogy between an intensional case and a strong liar, on which Skyrms draws, is as follows.[70] That the sum of 7 and 2 is odd is necessary, that the number of planets is odd is not—even though the sum of 7 and 2 *is* the number of planets; just so, the sentence on line 2 is true, the sentence on line 1 is not—even though the sentence on line 2 *is* the sentence on line 1. Skyrms proposes to solve the strong liar by construing 'true' as intensional. Truth values are to be assigned to sentence types *under descriptions*. Under the description 'the sentence on line 2', the sentence is true; under the description 'the sentence on line 1', it is not.

Formally, Skyrms proposes to assign truth values to sentence-description pairs (s,d), where d is a description of the sentence type s, and to do so in a way analogous to Kripke's fixed-point construction. He shows that, given a partial valuation of sentence-description pairs (which may distinguish between pairs that differ only in the description coordinate), a variant of Kripke's construction can be used to extend it. Skyrms's construction falls short of the desired goal since, being based on (the analogs of) Kripke's rules, it does not have the gap and jump rules, which are crucial. It works as long as we are moving through standard values; e.g., if we have already assigned (s, d) the value T, we can go on and assign (s', d') the value F, where s' is the negation of s and d' is some appropriate description obtained from d. But it does not provide a way for assigning T to the token on line 2 (i.e., to the pair (s, d), where s is the sentence and d is the description 'the sentence on line 2'). In general, it falls short whenever the jump rule is invoked.

---

[70] I use our two-line puzzle, the difference between it and the example used in Skyrms does not matter here.

Now, if desired, one can regard the pair (*s*,*d*) as a pointer to *s*. And, vice versa, a pointer to *s* can be changed to a pair (*s*, *d*) where *d* is some description derived from the pointer. *Formally*, my algorithm can be viewed as providing a truth-value assignment to sentences under descriptions, achieving thereby Skyrms's goal.[71] I do not see, however, how the failure of substitutivity arising in the strong-liar puzzle can be accounted by intensionality, without stretching "intensionality" so as to render it uninteresting. For, by the same token, one could subsume under it all other context-dependent phenomena, including, for example, indexicality. When each of Jack and Jill says, "I am tall," but only Jack is tall, we could explain the truth-value difference by the intensionality of 'true', namely, 'I am tall' is true, under the description 'the sentence uttered by Jack', but not true under the description 'the sentence uttered by Jill'. I doubt that we want to adopt this picture.

Intensionality is affected by the way in which referring terms pick their references. The sum of 7 and 2 must be the number 9, but it is only a cosmic accident, or so we think, that 9 is also the number of planets. But nothing in the liar situation hinges on some unclarity in the referential picture. The intensional effect disappears if the referential connection is made explicit. Given that the number of planets is the sum of 7 and 2, it *is* necessary that the number of planets is odd. Or, given that George IV knew that the author of *Waverly* was Scott, it is inconceivable that he wondered whether Scott wrote *Waverly*. But can we say that, given that the sentence type on line 2 is the sentence type on line 1, the sentence on line 1 is true? Certainly not.

Viewed in retrospect, both conceptions of truth—as an indexical, and as an intensional notion—are attempts to account for phenomena arising out of liar-like paradoxes by using tools that fall essentially within the framework of traditional logic. Both come short of providing a solution, but for different reasons.

HAIM GAIFMAN

Columbia University

---

[71] Indeed, judging from discussions I had with him, Skyrms regards the algorithm as a way of constructing the desired intensional fixed point.

IS IT REASONABLE TO REGRET THINGS ONE DID?*

We all know regret. Whether we take ourselves to be moral persons or immoralists, and, if the former, whether we have an Aristotelian, Kantian, or utilitarian understanding of morality (or no clear understanding at all), sometimes we regret what we did. The object of the following is not to demarcate what it is to regret things one did, to describe how it feels, or to state the precise conditions under which regret does and does not appear.[1] A rough explanation should suffice: regret is a painful feeling about something we did which we think was bad. More specific kinds of phenomena falling under this explanation, like remorse, repentance, feeling of guilt, will not be discussed in particular. Nor is the diversity of reasons for thinking it was bad of interest here: one can regret a purchase because it was too expensive, one can regret having behaved incorrectly because it diminishes one's standing, one can regret a mean remark because it hurt the other person; though it is the last kind of case, regret over something taken to be wrong for moral reasons, which will be in the foreground in what follows. The question is this: Given that we did something bad, is it reasonable for us to regret that we did it? The word 'reasonable' is to be understood here in its everyday sense, roughly equivalent to phrases like 'it's a good idea', 'it makes sense', 'it is recommendable'. Thus, the question can also be expressed as follows: Should we regret bad things we did?

It will be wondered whether the question itself is a reasonable one. After all, you feel regret, and what you feel is not up to you. To ask whether something is reasonable, however, makes sense only if it is up to you to do it or not. So the question whether regret is reasonable is mistaken right from the start. You might as well ask whether it is reasonable to have a headache.

This raises a large problem: How is what we feel related to the rest of our mental life, and how is it related to action, in particular to

---

* For helpful criticism of previous versions of this text I am grateful to Celène Abramson, Thomas Pogge, Amélie O Rorty, Nancy Schauber, and an audience at the University of Bielefeld, Germany.

[1] See Amélie O. Rorty, "Agent Regret," in Rorty, ed , *Explaining Emotions* (Berkeley: California UP, 1980), pp. 489–506; Gabriele Taylor, "Guilt and Remorse," in Taylor, *Pride, Shame, and Guilt: Emotions of Self-assessment* (New York: Oxford, 1985), pp. 85–107.

0022-362X/92/8905/262–73

moral action?[2] Without addressing this large question, a preliminary answer to the objection at hand may be given as follows. To ask whether something is reasonable makes sense not only in those cases where doing it or not is up to you in the sense that you can do it right on the spot. It is true that in this sense it is not up to us to regret things we did. We cannot at any given moment turn off the regret we feel about something we did, nor can we turn on a regret we do not feel. Yet over longer periods of time we learn to feel a certain way, and we learn not to feel a certain way any more, on this or that sort of occasion. This holds for individuals and for groups of people. For example, a feeling of reverence has passed out of use in my life. I did feel reverence for some people when I was much younger. Now I do not; and it seems this is due, not to a scarcity of suitable objects, but to my having outgrown this feeling altogether. I have lost this language; I do not just fail to find people with whom to speak it. Similarly, over the last few centuries people in the Western world have lost the feeling of awe. It still has a prominent place in Hobbes (and there is no reason to think that he is archaic on purpose here), but it has dried out since. The massive violence of the twentieth-century state inspires terror, not awe: there is nothing sacred about it. Now, it would be a long story, both in the individual and in the larger case, to explain how it happened that these feelings fell into desuetude. But it is very unlikely that a convincing story could be told, in either case, without referring to what people came to think, to see, to recognize in the process. People do not learn, or unlearn, to feel something *just* by being pushed around; it is also a matter of what they think about what happens to them.[3] For example, it was precisely Hobbes' own construction of an awe-producing political mechanism which, among numerous other influences, contributed to awe's becoming obsolete. And it is clear that I did not lose reverence for people like a key from my pocket. It was part of a whole process in which I came to have a different view of people who were important to me, and of myself, too. So in this sense some of our feelings are indeed up to us: whether some kind of feeling is available in a situation or not depends on what we have become and how we see things. In particular, then, it will depend, among other things, on whether we judge that feeling to be reasonable or not. Again, it is not that a feeling judged to be unreasonable is thereby and in that moment annihilated.[4] What sometimes happens is rather that a feeling judged to be unreasonable slowly recedes and eventually dies out, just because it makes no more sense to us. To be sure, this happens only sometimes: some feelings live on stubbornly in the face of better judgment. But sometimes they yield. Feeling is not in principle insensitive to insight. Perhaps it is here that the moral philosopher's task lies. He is not particularly effective in reforming action right on the spot in accordance with better insight. But he may, in the longer term, help people in learning to find sense, or to find no sense, in what they feel. Thus he may help in reaching a more sensible way of feeling and so in the end, perhaps, a more sensible way to act.

---

[2] There is a large literature on these topics. To mention only a few titles· Rorty, *op. cit.;* William Lyons, *Emotion* (New York: Cambridge, 1980), Ronald de Sousa, *Rationality and the Emotions* (Cambridge: MIT, 1987); Patricia S. Greenspan, *Emotions and Reasons: An Inquiry into Emotional Justification* (New York. 1988).

[3] The point here is the positive version of Epictet's famous line that what confuses people are not the things, but what they think about the things.

Spinoza raised the suspicion in his *Ethica* that to regret what one did is not reasonable: "Repentance is not a virtue, i.e. it does not arise from reason. Rather, he who repents what he did is twice miserable, i.e. impotent" (pt. IV, prop. 54). It is not the claim expressed in this sentence, however, which is under discussion here. That Spinoza speaks about repentance rather than regret does not make an important difference: his point, if he has one, applies to the latter phenomenon as well. The important difference lies in the concept of reason. For Spinoza, reason consists in the possession of adequate ideas about how things are.[5] This is a much narrower concept than the everyday notion indicated above. His claim that repentance, or regret, is not reasonable thus becomes uninterestingly weak. That we do not feel regret insofar as we entertain adequate ideas about ourselves and people and things around us, that is, insofar as we consider everything from the viewpoint of eternity (*ibid.*, prop. 44, corollary 2). is easy to grant. The interesting question is whether regret has a place in the broader field of what a reasonable human being does and feels, with 'reasonable' being understood in its ordinary sense. Spinoza's proposition does not address this question, but invites the reader to ask it.

Moreover, Spinoza suggests an argument for a negative answer. 'To regret what one did is to be twice miserable' means: a person is first miserable because he did something bad, and he is miserable a second time because he is in pain over what he did. 'To be miserable' must be understood here in an objective sense, not as 'feeling bad', but as 'being in a bad way'. This is so because the first misery is not a case of feeling bad. It is not, for otherwise it could not be distinguished from the second one. Rather, feeling bad is one kind of being in a bad way, which is why both the first and the second misery can indeed be called "misery." (Spinoza hints at this in adding the phrase 'i.e. impotent'.) What is under discussion is the state of an organism that is objectively 'low'. Still, one way for an organism to be low is to feel low. So what Spinoza is saying is this. The person who regrets what he did is in a bad way twice. Whoever did something bad is thereby, independently of how he feels about it, badly off. The pain he feels about what he did is for him to be badly off a second time. Thus, Spinoza's argument is: it is not reasonable, because one did something bad, to go and make things worse. But that is what regret is, double misery, the second for the sake of the first. So, regret is not reasonable. The same argument, incidentally, occurs in Nietzsche: "Never give room to repentance, but tell yourself at once: this would mean to add a second piece of stupidity to the first."[6]

---

[4] Hume seems to be supposing something as strong as this. "The moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition"; *A Treatise of Human Nature*, L. Selby-Bigge and P. Nidditch, eds (New York: Oxford, 1978), p. 416.

[5] *Ethica*, Pt. II, prop. 40, scholium 2

The argument also gives a clearer idea of the nonregretting person Spinoza has in view. It is not because this person never goes wrong or never admits to having gone wrong that he does not regret. Accordingly, Spinoza is not uselessly advising never to have grounds for regret. The reasonable person without regret has done something bad, and for the sake of the present argument it should be supposed that we all agree with him on this. So he is not one of those famous people who decline to regret because they do not recognize the law under which they appear guilty, from Don Giovanni's "No!" when the Commendatore asks him to repent, to Edith Piaf's "Non, je ne regrette rien." Unlike them, Spinoza's reasonable person is indeed once miserable. He is not immoral, nor is he bound to be heartless. He sees that what he did was wrong and he may be perfectly aware of the suffering he inflicted on others. He just does not grieve. Thomas Aquinas[7] held that the saints in the new life after judgment will be displeased by their earlier sins, but "without any sadness." These are sensible saints, Spinoza is saying, let us be like them, right now in this life.

---

[6] *The Wanderer and His Shadow,* 323. A similar sentiment is expressed in Ralph Waldo Emerson's essay "Self-reliance," as quoted by Rorty in "Agent Regret," p. 501.

[7] *Summa Theologiae,* III, qu. 84, art. 8

The opposite view has been put forward by Bernard Williams:[8] "it would be an insane concept of rationality" which insisted that a rational person would never experience sentiments of agent regret (*ibid.*, p. 29). In fact, here and in other passages Williams has a stronger claim in mind than the one under discussion here, and what he says is not specifically directed against the Spinozistic view just sketched. Williams holds that regret is sometimes appropriate not only with regard to intentional actions, but also with regard to consequences of action which the agent did not intend and could not even foresee, so that the field of potential regret extends to "almost anything for which one was causally responsible in virtue of something one intentionally did" (*ibid.*, p. 28). Williams holds this partly because it allows him a respectable concept of tragedy. It is true that tragedy presents us with extreme cases.[9] Still, what is central to tragedy is central to our life, too: the acceptance of things that just happened as part of what one has done and, thereby, of what one is.[10] Neither accident nor the machinations of the gods lead Oedipus into disaster, but rather what he did. It is what he did, for in the important sense "it was him," even though he did it unwittingly and unwillingly. This, evidently, is a far-reaching and quite doubtful thesis. Williams asks rhetorically: "Could we have, and do we want, a concept of agency by which what Oedipus said [namely, that he did not do it] would be simply true, and by which he would be seeing things rightly if for him it was straight off as though he had no part in it?" (*ibid.*, p. 30 fn.) But that is just to frighten the children. The answer to this question should simply be, "yes." Agency requires intentionality. In any case, what is under discussion here is not the strong thesis that regret is in order even with respect to unintended consequences of actions, but the weaker one that there are some cases where regretting what one did is reasonable. Given Spinoza's challenge, the thesis needs reasons.

Williams does not state his reasons explicitly, so an argument has to be constructed. The most natural consideration may be this: to regret something bad that one did is reasonable because it makes it more probable that one will do better in the future. There may be some hints in this direction in Williams's texts,[11] but it does not seem to be the reason on which his claim is based. Nor would it be a good reason. Regret does not in fact make doing better in the future more probable. The impression that it does arises from a confusion of Spinoza's reasonable person who does not regret with an immoral or heartless one. No doubt somebody who does not see anything wrong in what he did or has no sense for the suffering he caused is likely to do similar things again. That is not the case of the reasonable person who does not regret, however. He understands what he did, he knows that it was bad, it just does not pain him. And there is no reason to assume that it is simply being pained by what one did which makes doing better in the future more likely. Failure to learn from what went wrong comes in many forms. In moral matters there is the self-complacency of feeling terrible about what one did and leaving it at that, as well as the impassiveness of going on with business as usual. There are cases in all four fields of the two distinctions: there is regret with improving and without improving, and there is absence of regret with improving and without improving. That the ratio of improving to not improving is higher under regret than under absence of regret is not a plausible or justified assumption.

---

[8] "Moral Luck," in *Moral Luck* (New York: Cambridge, 1981)

[9] Williams, "Ethical Consistency," in *Problems of the Self* (New York: Cambridge, 1973), p. 173.

[10] Williams, "Moral Luck," pp 29f.

Second, regret could be defended with the idea of atonement. He who did evil must suffer evil, otherwise things will not return to order. Since the normal course of the world cannot be relied upon to make this happen, it is a good thing for the agent himself to pay for what he did by grieving about it. If he only registers that what he did was bad and even does everything he can to prevent his doing it again, he seems to get off the hook too easily. This leads to the idea of a regret that cleanses the heart: sincere grief can balance, and thereby remove, the agent's guilt. In this way, regret promises liberation. Suffering through what he did, the agent becomes free of the burden of the deed and can start afresh.

This defense has no chance to succeed, nor is there any indication that Williams would want to use it. Atonement is itself not a reasonable idea and cannot show regret to be reasonable. That he who did evil should suffer evil is neither required nor even desirable. However deeply it may have been inculcated in us that it ought to be the case that the good be happy and the evil suffer, the truth is that, if anything, it ought to be the case that everybody, good and evil alike, be happy. The suffering of him who did evil is only a further evil; it does nothing to remove the first. That the agent should, if possible, repair the damage he caused or recompense for it is another matter. Atonement is taken to eliminate, not the damage, but the guilt incurred in causing the damage. But this is a mythical idea. That we did what is bad cannot be undone. The promise of liberation through regret is a false promise. We should free ourselves, not of the burden of the deed, but of the idea that the deed is a burden.

---

[11] "Ethical Consistency," p. 175; "Politics and Moral Character," in *Moral Luck,* p. 62.

The arguments considered so far have parallels in the discussion of punishment, in the attack on punishment as a mere duplication of suffering, in the defense of punishment as a way to make betterment more probable, in the defense of punishment as atonement. Not so the following consideration, which probably comes closest to being the argument on which Williams's thesis is based. Having made the point, as quoted several paragraphs back, that a concept of rationality which excludes regret over what one did would be an insane one, Williams continues as follows:

> To insist on such a conception of rationality, moreover, would, apart from other kinds of absurdity, suggest a large falsehood: that we might, if we conducted ourselves clear-headedly enough, entirely detach ourselves from the unintentional aspects of our actions, relegating their costs to, so to speak, the insurance fund, and yet still retain our identity and character as agents.[12]

This statement is meant to support Williams's stronger thesis that regret is appropriate even with regard to unintended consequences of actions. The reasoning is: not to regret the unintended consequences of actions is to detach oneself from them, but in detaching oneself from them one fails to retain one's identity and character as an agent. The reasoning should apply with all the more force to the core area of action, intentional action, which is what is under discussion here. Not to regret what one did, Williams is saying, is to detach oneself from one's actions, but that is to lose one's identity and character as an agent.

So the reasoning is not: failure to regret would amount to a detachment from what one did, but that is impossible, so it makes no sense to try. Detachment is possible. What is wrong with it, according to Williams, is only that it is too costly; morally costly, that is. You pay detachment with your identity and character as an agent.

---

[12] "Moral Luck," p. 29.

The present argument, then, like the previous ones, tries to defend regret by indicating a benefit of it which is overlooked in Spinoza's challenge. Regret, the first argument said, makes better action in the future more likely. That did not appear plausible. Regret, the second argument said, obliterates the wrongdoing. That appeared false. Regret, now the third argument is saying, allows us to retain our identity and character as agents. This is unclear. We need to know three things. What is meant by 'retaining our identity and character as agents'? Why is it desirable to retain them? And why does it take regret to retain them?

The expression 'our identity and character as agents' cannot mean "our character of being agents." On this reading, Williams would claim that by not regretting one ceases to be an agent altogether, which is quite implausible. On the other hand, 'our identity and character as agents' cannot mean what is normally simply called "character," our individual qualities as they are manifested in how we act. On this reading, Williams would claim that failure to regret leads to the loss of one's character traits. That again is implausible. A generous, courageous, or short-tempered person does not stop being any of these things just by not regretting what he did. Rather, what is meant must be something like this. We are the agents we are not just by having done what we did. We are the agents we are by accepting these doings as ours. This is what the person who does not regret fails to do. He sheds what he did rather than identifying with it. He steals away out of his own deed. He does not doubt that he did it; rather, the point is that he does not positively take it on as his. What he thereby loses is after all, in one sense and to a certain extent, his character of being an agent. He loses it not in the sense that henceforth he is barred from agency. He loses it with respect to the past. He loses it in the sense that something he actually did gets cut out from what he is as an agent, given that what he is as an agent depends in part on what he accepts as his. "Where It was, there I shall become," Freud wrote. Where I was to be, there It is—this is what failure to retain one's identity and character as an agent means. It is an agent's crippling himself as an agent, by surrendering something he did to the realm of what just happens through him, rather than incorporating it in what he is as an agent. So what Williams is saying about the person who does not regret is that, in a specific sense, he lacks character. Not that he is characterless, i.e., bland. He lacks character the way a person does who, on encountering any sort of resistance, abandons people and projects to which he was committed: the person without regret abandons his deeds.

In short, Williams is Polonius: "This above all; to thine ownself be true."[13]

Why, really? This is the second of the questions distinguished above, about the desirability of retaining your identity and character as an agent. If failure to identify with what you did is held to make lack of regret morally costly, indeed on Williams's view prohibitively costly, it needs to be explained what is good about identifying with what you did. Why not throw your identity and character, in this sense, to the winds? There seems to be only one answer. You should identify with what you did because you did it after all. It is the truth about you that you realize in acknowledging agency. To thine own self, expressed in your actions, be true, for that is what you really are.

This is not a good answer, given the terms of the question. If your identity and character as an agent is something you may retain or fail to retain, then the option of retaining cannot be recommended on the grounds that thereby you acknowledge the truth about you. About this purported truth you are just called upon to decide, or else the question about retaining or not does not arise. If you are in the position to retain or not, then you are not in the position to stay true to yourself by either retaining or not retaining. "You must become who you are," Nietzsche said,[14] but the pathos of truth to oneself rests on an assumption, namely, that there is such a self to be true to, which it precisely undercuts, namely, by understanding truth to oneself as merely one option among others.[15] Thus, Williams's warning that by not regretting we fail to retain our identity and character as agents turns out to be another case of frightening the children. There is nothing detectably wrong with a failure to retain one's identity and character as an agent, in the sense of failing to identify with what one did. There is nothing detectably wrong with simply walking away from it. To be sure, we are not talking here about denying that one did it, in the ordinary sense. The person without regret and his opponent are agreed on the facts of what happened. The latter only claims that there is, beyond the trivial knowing that one was the doer of this deed, an acceptance of one's having done it as part of what one is, as part of one's identity, as this is often expressed;[16] and that such acceptance is what is required of agents. This claim is what is unfounded. We can live, and perhaps live better, without identity and character in this sense. Sure, we will have identity and character in the ordinary sense. We will be the individual agents we are, with these particular features manifesting themselves in action. The fact that we have identity and character in this sense, however, lends no support to the demand expressly to retain identity and character, in the special sense, through identification with what we did. Identity and character in the ordinary sense we will have in any case, so we may identify or not with what we did as we please. And sometimes it will make sense not to identify, and to say: "True, I did that, I even did it intentionally. But I am no more in it. You cannot find me there. That is history. It does not cast a shadow over what I do now." No doubt, somebody talking like that may be deceived and may, by anxiously denying the shadow, rather confirm its presence. The point is that this need not be the case. He may truly be free.

---

[13] William Shakespeare, Hamlet, I 3.

[14] *The Gay Science,* 270.

[15] Alexander Nehamas's discussion of Nietzsche's idea does not resolve this paradox; see his *Nietzsche: Life as Literature* (Cambridge: Harvard, 1985), ch. 6.

Suppose, for the sake of the argument, that all of that is wrong. Suppose, that is, that there is a sense of "our identity and character as agents" which makes the demand to retain our identity and character neither unjustified nor superfluous. That leaves us still without an answer to the third question, namely, why the retaining needs regret about the bad things one did. Why should not remembrance and understanding suffice, remembrance of what was done and understanding of what was bad about it? Why grieve in addition?

The best answer in defense of regret may be this: there is no understanding that the thing was wrong without the grief about having done it. This is not the cynical wisdom that "they won't remember what hasn't been beaten into them." The point is: to know that one did it and that it was bad is by itself painful. Spinoza may be right in saying that he who regrets is twice miserable. What he overlooks is the fact that the second misery is linked to the experience of the first. Thus, the only alternative to regret is simply to be blind to one's failings, and that this is not a desirable attitude is obvious. "Suffering we recognize that we did wrong," Antigone says.[17] The thought here is: only in suffering we recognize it. This is not evident.

---

[16] See again Williams, "Moral Luck," pp. 29f.: "if one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for, one must accept much that makes its claim on that sense solely in virtue of its being actual "

[17] Sophocles, *Antigone*, l. 926.

It is not evident that one could not see, in full clarity but without grief, what one did wrong. In fact, the contrary may be true: that with grief one could not see in full clarity what one did wrong. After all, to see such a thing is itself an achievement and, hurt, torn, and dejected by regret, we may not be capable of doing it. However this may be, the case for the necessity of regret for recognizing one's failing has not been made.

"Aren't you really returning now," it may be objected, "to Spinoza's understanding of reason, instead of staying with the ordinary one, as promised? Full clarity about what one did wrong, unclouded by grief, looks suspiciously like considering things from the point of view of eternity, not like something which it is a good idea to reach." Maybe this is what Spinoza meant by a consideration of things from the point of view of eternity: without pain just to get clear about matters. (Though 'eternity' speaks against this reading and suggests, rather, that this point of view lies beyond the human field.) If he meant that, then it is true, we are back with Spinoza's understanding of reason here. But it is not true that this is to quit the ordinary one. To look things straight in the face, unburdened by grief, is just a very good idea, and a good idea in the same sense in which it is a good idea, say, to have the snow tires mounted before the first snowfall. True, there is this difference: some people do mount their snow tires in time, but few people are likely to take in their failings without regret. Still, this shows only that some forms of unreason are more common and harder to eradicate than others. As discussed earlier, it does not show that this is a form of unreason we are stuck with once and for all.

"Your argument proves too much," it will be said. "This reasoning can be applied not only to regret of things done, but also to regret of things lost and above all to grief and mourning over people who are no longer with us, whether they died or went away. Let us not quarrel about whether it is possible to do without these emotions. Let us grant your point for now that a suitable *éducation sentimentale* may in the end eliminate them. But is that a desirable state? Imagine a man saying: 'Yes, it is a bad thing, we broke up, even though I had loved her from my heart; but it does not hurt me, I am alive and well as ever'. Isn't this a terribly cold, cruel, inhuman reaction? Something must be wrong with an idea of being reasonable which recommends such brutishness." There need not be anything cruel or brutish here. True, we often take such a reaction rightly as evidence that he did not love her so much after all or that, if he did, he now has lost, or indeed destroyed, his sense for why he loved her. But it need not be so. We may know independently that he did love her deeply and that his present reaction is not owed to his being blind to her. Also, a man speaking like this will often be speaking falsely, either hiding his pain from others or deceived about it himself. Again, this need not be so, however. And if everything is as he says it is: love, loss, and no wound, then so much the better. ("If you were the woman with whom he broke up, would you still say that?" Yes, especially then, I hope.) We do grieve about separations, but there is nothing recommendable or praiseworthy in grieving. If you can do without it, do not think you rather ought to grieve.

Or somebody may argue *ad hominem:* "I gather you are German. So obviously you are engaged in a little whitewash. Stripped of the philosophical disguise, your message is that Germans should stop doing penitence for what they did under the Nazis. We should all forget about the past, wipe the slate clean and turn to the business at hand as if nothing had happened. Alexander and Margarete Mitscherlich[18] lamented an inability to mourn, as they called it, in the German attitude to their past. You just give that defect a good name." There is a misunderstanding here. I am saying that penitence, in the sense of grieving over what one did, and mourning are not recommendable attitudes. I am saying that the Nazi who did the most horrendous things would still be wise to try, undejected by what he did, to do what now is right. To speak against penitence and mourning, however, is not to say that the events which penitence and mourning are about are any less bad. Nor is it to recommend forgetting them. On the contrary, it is with mourning and regret left behind that we may come to see what we did in the sharpest light. We should not suffer from what we did, we should face it.

RÜDIGER BITTNER

University of Bielefeld

---

[18] Alexander and Margarete Mitscherlich, *Die Unfähigkeit zu trauern* (München: Piper, 1967)

NOTES AND NEWS

The Department of Philosophy at Rice University is pleased to announce the appointment of George Sher as the Herbert S. Autrey Professor of Humanities.

The Department of Philosophy at Virginia Commonwealth University is pleased to announce the appointment of Eugene Mills as Assistant Professor.

The Department of Philosophy at the University of Waterloo is pleased to announce the appointment of Paul Thagard as Professor

Joan C. Callahan has been elected Executive Secretary of the Society for Philosophy and Public Affairs. Please address all correspondence with the Society to Prof. Callahan, Phil. Dept., Univ. of Kentucky, Lexington, KY 40606-0027.

The National Council for Critical Analysis seeks a person or institution willing to take over *The Journal of Critical Analysis* and/or *The Journal of Pre-College Philosophy*. Anyone interested should communicate with Dr. Pasqual S. Schievella, NCCA, P.O. Box 137, Port Jefferson, NY 11777.

Penn State Press announces a new series, Re-Reading the Canon, with Nancy Tuana as General Editor. This new series will consist of edited collections of essays, some original and some previously published, offering feminist reinterpretations of the writings of major figures in the Western philosophical tradition. Devoted to the work of a single philosopher, each volume will contain essays covering the full range of the philosopher's thought and representing the diversity of approaches now being used by feminist critics. The editor of each volume will write an introduction and prepare an extensive bibliography, in addition to selecting previously published essays and commissioning new contributions. The series will begin with a volume on Plato edited by Nancy Tuana. Other early volumes will focus on Aristotle, Locke, Marx, Wittgenstein, De Beauvoir, Foucault, and Derrida. Inquiries should be directed to Prof. Nancy Tuana, School of Arts and Humanities, Univ. of Texas/Dallas, Box 830688, Richardson, TX 75083-0688.

The Department of Philosophy at the Virginia Polytechnic Institute and State University is pleased to announce that it will be offering a course of study leading to the M.A. in Philosophy, beginning in the Fall. Special areas of concentration include History of Philosophy, Philosophy of Science, and Ethics. The Univ. and Dept take affirmative action very seriously. We are especially interested in receiving applications from women and people of color. Please direct inquiries to Graduate Program, Phil. Dept., Virginia Tech & State Univ., Blacksburg, VA 24061-0126.

TIME, COMMUNICATION AND KNOWLEDGE

FORTHCOMING

GOD, ETERNITY AND THE NATURE OF TIME

Alan G. Padgett

This book provides a thorough analysis and discussion of the nature of time and eternity, focusing on the timelessness of God. After offering both a historical and biblical perspective of the doctrine of eternity, Padgett rejects both the "everlasting" and "timeless" theory of eternity, arguing for a third, mediating view of God as in some sense temporal, yet also "relatively" timeless.

June • 208 pp. • ISBN 0-312-06813-1 • $55.00

THE KNOWLEDGE THAT ENDURES

Coleridge, German Philosophy and the Logic of Romantic Thought

Gerald McNiece

1991 • 244 pp. • $55.00

ISBN 0-312-06799-2

ALSO FORTHCOMING

COMMUNICATIVE ETHICS IN THEORY AND PRACTICE

Niels Thomassen;

Translated by John Irons

July • 240 pp. • $49.95

ISBN 0-312-06806-9

KIERKEGAARD ON ART AND COMMUNICATIONS

Edited by George Pattison

August • 192 pp. • $49.95

ISBN 0-312-07478-6

ST. MARTIN'S PRESS • Scholarly & Reference

175 Fifth Ave. • New York, NY 10010 • 1-800-221-7945

THE JOURNAL OF PHILOSOPHY

is pleased to announce the publication of its

TWENTY-FIVE YEAR INDEX, 1964–1988

VOLUME LXI TO VOLUME LXXXV

AUTHORS, SUBJECTS, & REVIEWS

Supplement to its

FIFTY-YEAR INDEX, 1904–1953

TEN-YEAR SUPPLEMENT, 1954–1963

It is now available at special prices

25-year (1964–1988) $25 00 cloth/$10 00 paper

also available

50-year (1904–1953) $12 00 cloth 10-year (1954–1963) $3 50 paper

Shipping $5 00 (first)/$2 50 (each addl)

Please send a check or money order to

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through August 31, 1992 Orders and payment must be received by then Paper copies will be mailed immediately, but cloth copies will take 6–8 weeks for delivery

CAMBRIDGE PRESENTS...

Essays on Henry Sidgwick

Bart Schultz, Editor

In this volume, a distinguished group of philosophers reassess the full range of Sidgwick's work, not simply his ethical theory but also his contributions as a historian of philosophy, a political theorist, and a reformer.

Contributors: *David O. Brink, Thomas Christiano, Stefan Collini, John Deigh, Alan Donagen, William K. Frankena, Russell Hardin, T. H. Irwin, James T. Kloppenberg, J. L. Mackie, J. B. Schneewind, Marcus G. Singer, Nicholas P. White*

39151-2 Hardcover $65.00

The Philosophical Writings of Descartes

Volume 3: The Correspondence

René Descartes

John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny, Editors

Volume III contains 207 of Descartes' letters, over half of which have previously not been translated into English and incorporates Anthony Kenny's celebrated translation of selected philosophical letters, first published in 1970.

40323-5 Hardcover $65 00
42350-3 Paperback $19.95

The Cambridge History of Political Thought 1450–1700

J. H. Burns, Editor
With Mark Goldie

This book presents a comprehensive account of the development of European political thinking through the Renaissance to the seventeenth century.

24716-0 Hardcover $120.00

Now in paperback...

Trials and Punishments

R. A. Duff

Cambridge Studies in Philosophy

40761-3 Paperback $16.95

Nietzsche contra Rousseau

A Study of Nietzsche's Moral and Political Thought

Keith Ansell-Pearson

This important contribution to Nietzche's political thought takes a serious look at Nietzsche as political thinker and relates his political ideas to the dominant traditions of modern political thought.

41173-4 Hardcover $49 95

The Rise of Neo-Kantianism

Klaus Christian Kohnke

R. J. Hollingdale, Translator

Foreword by Lewis White Beck

A major contribution to the intellectual history of the nineteenth century, Kohnke's book has profound implications for the way in which the history of philosophy is pursued.

Published with the support of the Exxon Education Foundation
Ideas in Context 20

37336-0 Hardcover $64 95

Theory of Societal Constitutionalism

Foundations of a Non-Marxist Critical Theory

David Sciulli

The author develops an alternative conceptual framework and explains how his theory can be used to assess whether social order in a society, whether democratic or authoritarian in political rule, is characterized by some degree of social authoritarianism.

ASA Rose Monograph Series

41040-1 Hardcover $59.95

Now in paperback...

The Cambridge History of Medieval Political Thought c. 350–c. 1450

J. H. Burns, Editor

24324-6 Hardcover $110.00
42388-0 Paperback $34 95

ESSENTIAL PHILOSOPHY

The Fixation of Belief and Its Undoing

Changing Beliefs Through Inquiry

Isaac Levi

Isaac Levi's new book is concerned with how one can justify changing one's beliefs. The discussion is deeply informed by the belief-doubt model advocated by C. S. Peirce and John Dewey, of which the book provides a substantial analysis.

41266-8 Hardcover $34 50

Paradoxes of Belief and Strategic Rationality

Robert C. Koons

"The approach is powerful, original, and very substantial from an intellectual point of view." —Tyler Burge, University of California, Los Angeles

Cambridge Studies in Probability, Induction and Decision Theory

41269-2 Hardcover $49.95

The Future of Folk Psychology

Intentionality and Cognitive Science

John D. Greenwood, Editor

The essays are concerned with our everyday and developed scientific systems of explanation of human behavior.

Contributors: *Jonathan Bennett, Simon Blackburn, Paul M. Churchland, Daniel C Dennett, J. A Fodor, Joseph Garon, John D. Greenwood, John Heil, Terence Horgan, Richard McDonough, Joseph Margolis, William Ramsey, Jay Rosenberg, Steven Stich, James Woodward.*

40335-9 Hardcover $44 50
40898-9 Paperback $12 95

The Metaphysics of Morals

Immanuel Kant

Mary Gregor, Translator

This is the only complete translation of Kant's major work in applied moral philosophy in which he deals with the basic principles of rights and of virtues.

Texts in German Philosophy

30372-9 Hardcover $49.50
31657-X Paperback $16.95

Matters of Metaphysics

D. H. Mellor

This selection of Professor Mellor's work gathers together sixteen major papers on related topics written over the past fifteen years to form a complete modern metaphysics.

41117-3 Hardcover $44 95

Substance and Predication in Aristotle

Frank A. Lewis

This book takes up the central themes of Aristotle's metaphysical theory and the various transformations they undergo prior to their full expression in the *Metaphysics*.

39159-8 Hardcover $44 50

The Disappearance of Time

An Essay on the Philosophy of Kurt Godel

Palle Yourgrau

This book explores the great logician Kurt Gödel's conception of time, existence, and truth with special reference to Plato, Aristotle, Kant, and Frege.

41012-6 Hardcover $39 50

CAMBRIDGE
UNIVERSITY PRESS

40 West 20th Street, New York, NY 10011-4211 Call toll-free 800-872-7423
MasterCard/VISA accepted Prices subject to change

CLASSIC PHILOSOPHERS

The Complete Essays

Michel de Montaigne

Translated by M. A. Screech

This triumph of literary scholarship is the first new English translation in more than thirty years of the complete work of the creator of the modern essay. In 1572, Michel de Montaigne (1533-1592) began to write about himself and his life as an ordinary man: one who read, traveled, endured war and peace, and attempted to find meaning in his experiences. His essays are invaluable as historical documents and as instances of a probing, skeptical intelligence trying to use itself as a lens for the entire human condition. 1344 pp.

Allen Lane hardcover 0-7139-9072-4 $60.00

MONTAIGNE AND MELANCHOLY

The Wisdom of the Essays

M. A. Screech

This remarkable book by renowned translator Screech presents "a sensitive probe into how Montaigne resolved for himself the age-old ambiguities of melancholia and, in doing so, spoke of what he called the 'human condition.' "—Roy Porter, *The London Review of Books*. 224 pp.

Penguin paperback 0-14-014362-9 $10.00

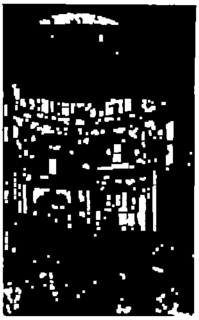
ESSENTIALS OF GOVERNMENT ETHICS

Edited by Peter Madsen, Ph.D. and Jay M. Shafritz, Ph.D.

Foreword by Alfred Blumenstein

From Plato and Machiavelli to such modern philosophers and politicians as Sissela Bok and Mario Cuomo, this is the first book to bring together key source readings on issues of government ethics, ranging from abstract arguments to detailed discussions. 512 pp.

Meridian paperback 0-452-01091-8 $15.00



CAPITAL
Volume 1
Karl Marx
Translated by Ben Fowkes
Introduction by Ernest Mandel

Marx's classic 1867 study of capitalistic production remains marvelously lucid in its vision of the relationship between labor and value, the role of money, and the conflict between the classes. 1152 pp.
Penguin Classic 0-14-044568-4 $14.95

THE ART OF RHETORIC
Aristotle
Translated and Introduced by Hugh-Lawson Tancred

By establishing the methods of informal reasoning, offering the first aesthetic evaluation of prose style, and exploring the detailed psychology of the emotions, Aristotle's rhetoric manual uncovers the very roots of persuasion itself. 304 pp.
Penguin Classic 0-14-044510-2 $8.95

EARLY WRITINGS
Karl Marx
Translated by Rodney Livingstone and Gregor Benton
Introduction by Lucio Colletti

This collection presents the foundations of the grand edifice of Marxism and a look at the early work of one of the greatest intellects in history. 464 pp.
Penguin Classic 0-14-044574-7 $9.95
Available July 1992

ATLAS SHRUGGED
35th Anniversary Edition
Ayn Rand
New Introduction by Leonard Peikoff

Rand's most successful novel embodies the Objectivist philosophy through the story of a man who vows to singlehandedly halt the inexorable progress of the world. 1084 pp.
Dutton hardcover 0-525-93418-9 $30.00
Signet paperback 0-451-17192-6 $6.99

WILLIAM JAMES
Writings 1878-1899
Edited by Gerald E. Myers

Often described as America's foremost philosopher, psychologist, and writer on religion, James' extensive and highly influential writings have placed him beside Emerson and Whitman as a classic exponent of American democratic culture. 1200 pp.
Library of America 0-940450-72-0 $35.00

THE LAWS OF MANU
Translated and Introduced by Wendy Doniger with Brian K. Smith

Written by several Brahmin hands from 200 BC to 200 A.D., this spiritual guide draws on jurisprudence, philosophy, and religion to create a model on which to base morality. 368 pp.
Penguin Classic 0-14-044540-4 $10.95

SELECTED WRITINGS
John Ruskin
Edited and Introduced by Kenneth Clark

This extensively annotated volume traces the work of John Ruskin (1819-1900), presenting a selection of his writings on art, literature, politics, nature, and his own character to convey the range of his intellect and a representation of the age for which he wrote. 384 pp.
Penguin Classic 0-14-043355-4 $10.95

PENGUIN USA
Academic Marketing Department
375 Hudson Street, New York, NY 10014-3657

NEW FROM
THE FREE PRESS

THE END OF HISTORY AND THE LAST MAN

Francis Fukuyama

“Fukuyama provides a fascinating historical and philosophical setting for the twenty-first century His discussion of the idea of *thymos* may prove to be even more important than his theory of the end of history.”
—**Tom Wolfe**

1992 0-02-910975-2 $24.95

MEANING IN LIFE
The Creation of Value

Irving Singer, *Massachusetts Institute of Technology*

“The nearest thing to a rational discernment of meaning in life, Irving Singer suggests, is the discernment of goals and values that we can find gratification and fulfillment in pursuing. He offers wise counsel to that end, along with rewarding glimpses into a wide range of literature.”
—**W.V. Quine**, Harvard University

1992 0-02-928905-X $19.95

THE SEVEN DEADLY SINS
Jewish, Classical, and Christian Reflections on Human Nature

Solomon Schimmel, *Hebrew College*

“This book mediates the moral wisdom of antiquity in a very useful and engaging way. It shows that the temptations to which we are all subject do not change. What does change is our willingness to recognize and overcome them. Few of us can afford not to read this volume.”
—**Jon D. Levenson**, Harvard University

1992 0-02-927901-1 $19.95

THE FINE LINE
Making Distinctions in Everyday Life

Eviatar Zerubavel, *Rutgers University*

“An eclectic and impressive compendium of mental boundaries, boundary shifts and shades of grey. A valuable and thought-provoking excursion into how the human mind works, and Zerubavel pulls off the difficult trick of writing in a scholarly and stimulating way at the same time It's well worth reading for people inside, outside, and ambiguously straddling the boundary of the new field of cognitive science.”
—**Douglas Hofstadter**, author of *Godel, Escher, Bach*

1991 0-02-934420-4 $22.95

For VISA, MasterCard or American Express orders,
call toll-free 1-800-323-7445

Fp **THE FREE PRESS**
A Division of Macmillan/A Maxwell Macmillan Company
866 Third Avenue, New York, NY 10022

The Libidinal Economy

Jean-François Lyotard

Translated by Iain Hamilton Grant

Lyotard's stunning *Économie libidinale* (1974)—a dense, brilliant piece which argues that political economy is in fact charged with passions, and reciprocally, that passions are infused with the political. *Theories of Contemporary Culture*

cloth $49.95 paper $17.95

Explorations in Feminist Ethics

Theory and Practice

Edited by Eve Browning Cole and Susan Coultrap-McQuin

A range of recent work in feminist ethics, exploring such issues as the ethic of care, Pythagoreanism, existentialism, utilitarianism, emotion, moral vision, and applying feminist ethics to work situations.

cloth $35.00 paper $12.95

Pandora and Occam

On the Limits of Language and Literature

Horst Ruthrof

Evoking Pandora and Occam as metaphoric corner posts in an argument about language as discourse, Ruthrof brings analytic philosophy to bear on issues of Continental philosophy. *Advances in Semiotics*

cloth $35.00

Cooking, Eating, Thinking

Transformative Philosophies of Food

Edited by Deane W. Curtin and Lisa M. Heldke

The collection offers new readings of a number of texts—religious, philosophical, anthropological, culinary, poetic, and economic—arguing the existence of a philosophy of food.

cloth $45.00 paper $19.95

Studies in Continental Thought

The Other Heading

Reflections on Today's Europe

Jacques Derrida

Translated by Pascale-Anne Brault and Michael B. Naas

Jacques Derrida on the identity and future of the "new Europe" (post-EEC, post-USSR) and on the outlook for democracy there.

cloth $17.95

Parmenides

Martin Heidegger

Translated by André Schuwer and Richard Rojcewicz

Parmenides, a lecture course delivered by Heidegger at the University of Freiburg in 1942-1943, presents a highly original interpretation of ancient Greek philosophy.

cloth $29.95

Reading Heidegger

Commemorations

Edited by John Sallis

Some of the world's most distinguished interpreters of Heidegger's philosophy discuss and extend his thought as it bears on a broad range of fundamental philosophical questions.

cloth $45.00 paper $19.95

History of the Concept of Time

Prolegomena

Martin Heidegger

Translated by Theodore Kisiel

". . . an excellent translation of an extremely important book." —*The Modern Schoolman*

Now in paperback! $17.50

Indiana
University Press

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709 Philosophy Hall, Columbia University
New York, New York 10027

Kant and the Exact Sciences

Michael Friedman

Kant sought throughout his life to provide a philosophy adequate to the sciences of his time—especially Euclidean geometry and Newtonian physics In this new book, Michael Friedman argues that Kant's continuing efforts to find a metaphysics that could provide a foundation for the sciences is of the utmost importance in understanding the development of his philosophical thought from its earliest beginnings in the thesis of 1747, through the *Critique of Pure Reason*, to his last unpublished writings in the *Opus postumum*

$45 00 cloth

Realism with a Human Face

Hilary Putnam

Edited and introduced by
James Conant

"There are many strengths in these essays Like all his work, they are highly readable, they are on deep and general issues, they contain forceful critiques, particularly of such diverse figures as Ayer and Rorty, and Putnam's first-order views in ethics are compassionate and imaginative any reader of these vivid and provocative essays will be forced to rethink his views somewhere"

—*Times Literary Supplement*

$15 95 paper

Sources of the Self

The Making of the
Modern Identity

Charles Taylor

"*Sources of the Self* is in every sense a large book in length and in the range of what it covers, but above all in the generosity and breadth of its sympathies and its interest in humanity Few books on such large subjects are so engaging"

—*New York Review of Books*

$18 95 paper

TO BE PUBLISHED IN
PAPERBACK IN FALL 1992

The Realm of Rights

Judith Jarvis Thomson

Wise Choices, Apt Feelings

A Theory of Normative
Judgement

Allan Gibbard

**Harvard
University
Press**

Cambridge, MA 02138
(617) 495-2480

Princeton

Concluding Unscientific Postscript to Philosophical Fragments

New Edition

Søren Kierkegaard

Edited and translated, with introduction and notes,
by Howard V. Hong and Edna H. Hong

Concluding Unscientific Postscript is on one level a philosophical jest, yet, on another it is a characterization of the subjective thinker's relation to the truth of Christianity.

Kierkegaard's Writings, XII

Volume I
Paper: $19.95 ISBN 0-691-02081-7
Cloth: $69.50 ISBN 0-691-07395-3

Volume II
Paper: $14.95 ISBN 0-691-02082-5
Cloth: $45.00 ISBN 0-691-07402-X

Set
Paper: $29.95 ISBN 0-691-02083-3
Cloth: $99.50 ISBN 0-691-07403-8

New in paperback

Eighteen Upbuilding Discourses

Søren Kierkegaard

Edited and translated, with introduction and notes,
by Howard V. Hong and Edna H. Hong

Kierkegaard's Writings, V

Now in paper: $17.95 ISBN 0-691-02087-6

New in paperback

The Chances of Explanation

Causal Explanation in the Social, Medical, and Physical Sciences

Paul Humphreys

This book provides a post-positivist theory of deterministic and probabilistic causality that supports both quantitative and qualitative explanations.

"... by far the best thing written on probabilistic causality and probabilistic explanation to date."—Wesley C. Salmon, University of Pittsburgh

Now in paper: $14.95 ISBN 0-691-02086-8

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540
ORDERS: 800-777-4726 • OR FROM YOUR LOCAL BOOKSTORE

12 AUG 1992

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 6
JUNE 1992

page
275 *Indispensability and Practice* Penelope Maddy
290 *The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction* Jaakko Hintikka and Gabriel Sandu

BOOK REVIEWS

316 Allen Wood: *Hegel's Ethical Thought* Frederick Neuhouser
321 Penelope Maddy: *Realism in Mathematics* Shaughan Lavine
326 Judith Jarvis Thomson: *The Realm of Rights* Carl Wellman
330 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 6, JUNE 1992

INDISPENSABILITY AND PRACTICE*

For some time now, philosophical thinking about mathematics has been profoundly influenced by arguments based on its applications in natural science, the so-called "indispensability arguments." The general idea traces back at least to Gottlob Frege,[1] but contemporary versions stem from the writings of W. V. Quine,[2] and later, Hilary Putnam.[3] Much contemporary philosophy of mathematics (including my own) operates within the parameters of the indispensability arguments; they are called upon to motivate various versions of nominalism, as well as to support various versions of realism. Still, attention to practice, both scientific and mathematical, has recently led me to doubt their efficacy. I shall try to explain these doubts in what follows. If they are legitimate, we will be forced to rethink much of current orthodoxy in the philosophy of mathematics.

I NATURALISTIC BACKGROUND

An argument based on scientific and mathematical practice can only succeed from a sufficiently naturalistic perspective. Familiar Quinean naturalism counsels us to reject prescientific first philosophy in favor of an approach that begins within our current best scientific theory:

---

* My thanks go to the NSF (DIR-9004168) and to the UC/Irvine Academic Senate Committee on Research for their support.

[1] ". . . it is applicability alone which elevates arithmetic from a game to the rank of a science," P. Geach and M. Black, eds., *Translations from the Philosophical Writings of Gottlob Frege* (Cambridge: Blackwell, 1970), p. 187. Although Frege speaks of arithmetic, my focus here will be on analysis, broadly construed

[2] See, e.g., "On What There Is," repr. in *From a Logical Point of View*, 2nd ed. (Cambridge: Harvard, 1980), pp. 1–19; and "Carnap and Logical Truth," repr. in *The Ways of Paradox*, rev. ed. (Cambridge: Harvard, 1976), pp. 107–32.

[3] See, e.g., "What Is Mathematical Truth?" and "Philosophy of Logic," repr. in *Mathematics, Matter and Method*, 2nd ed. (New York: Cambridge, 1979), pp. 60–78, 323–57.

0022-362X/92/8906/275–89

> . . . naturalism: abandonment of the goal of a first philosophy. It sees natural science as an inquiry into reality, fallible and corrigible but not answerable to any supra-scientific tribunal, and not in need of any justification beyond observation and the hypothetico-deductive method . . . The naturalistic philosopher begins his reasoning within the inherited world theory as a going concern. He tentatively believes all of it, but believes also that some unidentified portions are wrong. He tries to improve, clarify, and understand the system from within. He is the busy sailor adrift on Neurath's boat.[4]

Epistemological studies, in particular, are to be carried out within science, with the help of relevant psychological theories.

From the perspective of this scientific naturalism, a philosopher can criticize scientific practice, but only on scientific grounds, as a scientist might do, for good scientific reasons. This is enough to ratify an appeal to scientific practice in philosophical contexts: because scientific practice can only be questioned on scientific grounds, a conflict between scientific practice and philosophy must be resolved by revising the philosophy. So, for example, if scientific practice holds that p does or does not count as evidence for q, to disagree on philosophical grounds is an offense against naturalism.

As we shall see, however, it is not clear that Quine intends to extend this naturalistic faith in practice to the practice of mathematics. Leaving Quine aside for the moment, we must ask ourselves what the role of the philosophy of mathematics should be. Mathematics, after all, is an immensely successful enterprise in its own right, older, in fact, than experimental natural science. As such, it surely deserves a philosophical effort to understand it as practiced, as a going concern. Indeed, as in any discipline, there remain conceptual confusions in mathematics that might be clarified by philosophical analysis, providing that analysis is sensitive to the realities of actual mathematics. If it is to serve these purposes, a philosophical account of mathematics must not disregard the evidential relations of practice or recommend reforms on nonmathematical grounds.

These are, in my view, proper goals for the philosophy of mathematics. We, as philosophers of mathematics, should provide an account of mathematics as practiced, and we should make a contribution to unraveling the conceptual confusions of contemporary mathematics. So it is against this backdrop that I shall assess the indispensability arguments, or rather, the view of mathematics the indispensability arguments generate. They will be judged by their ability to account for actual mathematics as practiced.

---

[4] Quine, "Five Milestones of Empiricism," in *Theories and Things* (Cambridge: Harvard, 1981), pp. 67–72, here p 72

II INDISPENSABILITY

The original indispensability arguments were aimed at those who would draw a weighty ontological or epistemological distinction between natural science and mathematics. To those tempted to admit the existence of electrons while denying the existence of numbers, Quine[5] points out that

> Ordinary interpreted scientific discourse is as irredeemably committed to abstract objects—to nations, species, numbers, functions, sets—as it is to apples and other bodies. All these things figure as values of the variables in our overall system of the world. The numbers and functions contribute just as genuinely to physical theory as do hypothetical particles (*ibid.*, pp. 149–50).

But, Quine's opponent insists, the scientific hypotheses in our theory are tested by experiment, and the mathematical ones are not; surely the two can be distinguished on these grounds. To which Quine replies,

> The situation may seem to be saved, for ordinary hypotheses in natural science, by there being some indirect but eventual confrontation with empirical data. However, this confrontation can be remote; and, conversely, some such remote confrontation with experience may be claimed even for pure mathematics and elementary logic. The semblance of a difference in this respect is largely due to overemphasis of departmental boundaries. For a self-contained theory which we can check with experience includes, in point of fact, not only its various theoretical hypotheses of so-called natural science but also such portions of logic and mathematics as it makes use of.[6]

Thus, the applied mathematics is confirmed along with the physical theory in which it figures.

Of course, it is not enough for a piece of mathematics simply to appear in a confirmed scientific theory. For any theory T, there is another theory T' just like T except that T' posits a bunch of new particles designed to have no affect on the phenomena T predicts. Any experiment confirming T under these circumstances would also (in some sense) confirm T', but we do not take this as evidence for the existence of the new particles because they are "dispensable," i.e., there is an equally good, indeed better theory of the same phenomena, namely, T, that does not postulate them. The mathematical apparatus of modern physics does not seem to be dispensable in this way; indeed, Putnam has emphasized that many physical hypotheses cannot even been stated without reference to numbers, functions, etc.[7]

---

[5] "Success and Limits of Mathematization," in *Theories and Things*, pp. 148–55.

[6] "Carnap and Logical Truth," p. 121.

So a simple indispensability argument for the existence of mathematical entities goes like this: we have good reason to believe our best scientific theories, and mathematical entities are indispensable to those theories, so we have good reason to believe in mathematical entities. Mathematics is thus on an ontological par with natural science. Furthermore, the evidence that confirms scientific theories also confirms the required mathematics, so mathematics and natural science are on an epistemological par as well.

Unfortunately, there is a prima facie difficulty reconciling this view of mathematics with mathematical practice.[8] We are told we have good reason to believe in mathematical entities because they play an indispensable role in physical science, but what about mathematical entities that do not, at least to date, figure in applications? Some of these are admissible, Quine[9] tells us,

> . . . insofar as they come of a simplificatory rounding out, but anything further is on a par rather with uninterpreted systems (*ibid.*, p. 788)

So in particular,

> I recognize indenumerable infinites only because they are forced on me by the simplest known systematizations of more welcome matters. Magnitudes in excess of such demands, e.g. $\beth_\omega$ or inaccessible numbers, I look upon only as mathematical recreation and without ontological rights.[10]

The support of the simple indispensability argument extends to mathematical entities actually employed in science, and only a bit beyond.

The trouble is that this does not square with the actual mathematical attitude toward unapplied mathematics. Set theorists appeal to various sorts of nondemonstrative arguments in support of their customary axioms, and these logically imply the existence of $\beth_\omega$. Inaccessibles are not guaranteed by the axioms, but evidence is cited on their behalf nevertheless. If mathematics is understood purely on the basis of the simple indispensability argument, these mathematical evidential methods no longer count as legitimate supports; what matters is applicability alone. Here simple indispensability theory rejects accepted mathematical practices on nonmathematical grounds, thus ruling itself out as the desired philosophical account of mathematics as practiced.

---

[7] Hartry Field disputes this in his *Science without Numbers* (Princeton: University Press, 1980), and *Realism, Mathematics and Modality* (Cambridge: Blackwell, 1989), but the copious secondary literature remains unconvinced.

[8] Versions of this concern appear in C. Chihara, *Constructibility and Mathematical Existence* (New York: Oxford, 1990), p. 15; and in my *Realism in Mathematics* (New York: Oxford, 1990), pp. 30–1.

[9] "Review of Charles Parsons's *Mathematics in Philosophy*," this JOURNAL, LXXXI, 12 (December 1984): 783–94

[10] Quine, "Reply to Charles Parsons," in *The Philosophy of W. V. Quine*, L. Hahn and P. Schilpp, eds. (La Salle, IL: Open Court, 1986), pp. 396–403; here p. 400.

So simple indispensability will not do, if we are to remain faithful to mathematical practice. We insist on remaining faithful to mathematical practice because we earlier endorsed a brand of naturalism that includes mathematics. But it is worth noting that even a retreat to purely nonmathematical naturalism (forgetting our commitment to actual mathematical practice) will not entirely solve this problem. From the point of view of science-only naturalism, the applied part of mathematics is admitted as a part of science, as a legitimate plank in Neurath's boat; unapplied mathematics is ignored as unscientific. But even for applied mathematics there is a clash with practice. Mathematicians believe the theorems of number theory and analysis not to the extent that they are useful in applications but insofar as they are provable from the appropriate axioms. To support the adoption of these axioms, number theorists and analysts may appeal to mathematical intuition, or the elegant systematization of mathematical practice, or other intramathematical considerations, but they are unlikely to cite successful applications. So the trouble is not just that the simple indispensability argument shortchanges unapplied mathematics; it also misrepresents the methodological realities of the mathematics that is applied.

There is, however, a modified approach to indispensability considerations which gets around this difficulty. So far, on the simple approach, we have been assuming that the indispensability of (some) mathematical entities in well-confirmed natural science provides both the justification for admitting those mathematical things into our ontology and the proper methodology for their investigation. But perhaps these two—ontological justification and proper method—can be separated. We could argue, first, on the purely ontological front, that the successful application of mathematics gives us good reason to believe that there are mathematical things. Then, given that mathematical things exist, we ask: By what methods can we best determine precisely what mathematical things there are and what properties these things enjoy? To this, our experience to date resoundingly answers: by mathematical methods, the very methods mathematicians use; these methods have effectively produced all of mathematics, including the part so far applied in physical science.

From this point of view, a modified indispensability argument first guarantees that mathematics has a proper ontology, then endorses (in a tentative, naturalistic spirit) its actual methods for investigating that ontology. For example, the calculus is indispensable in physics; the set-theoretic continuum provides our best account of the calculus; indispensability thus justifies our belief in the set-theoretic continuum, and so, in the set-theoretic methods that generate it; examined and extended in mathematically justifiable ways, this yields Zermelo-Fraenkel set theory. Given its power, this modified indispensability theory of mathematics stands a good chance of squaring with practice, so it will be preferred in what follows.[11]

III. THE SCIENTIFIC PRACTICE OBJECTION

My first reservation about indispensability theory stems from some fairly commonplace observations about the practice of natural science, especially physics. The indispensability argument speaks of a scientific theory *T*, well-confirmed by appropriate means and seamless, all parts on an ontological and epistemic par. This seamlessness is essential to guaranteeing that empirical confirmation applies to the mathematics as well as the physics, or better, to the mathematized physics as well as the unmathematized physics. Quine's[12] vivid phrases are well-known: "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (*ibid.*, p. 41).

Logically speaking, this holistic doctrine is unassailable, but the actual practice of science presents a very different picture. Historically, we find a wide range of attitudes toward the components of well-confirmed theories, from belief to grudging tolerance to outright rejection. For example, though atomic theory was well-confirmed by almost any philosopher's standard as early as 1860, some scientists remained skeptical until the turn of the century—when certain ingenious experiments provided so-called "direct verification"—and even the supporters of atoms felt this early skepticism to be scientifically justified.[13] This is not to say that the skeptics necessarily recommended the removal of atoms from, say, chemical theory; they did, however, hold that only the directly verifiable consequences of atomic theory should be believed, whatever the explanatory power or the fruitfulness or the systematic advantages of thinking in terms of atoms. In other words, the confirmation provided by experimental success extended only so far into the atomic-based chemical theory *T*, not to the point of confirming its statements about the existence of atoms. This episode provides no comfort to the van Fraassenite, because the existence of atoms was eventually established, but it does show scientists requiring more of a theory than the sort of theoretical virtues typically discussed by philosophers.

---

[11] This is more or less the position of my *Realism in Mathematics*.

[12] "Two Dogmas of Empiricism," repr. in *From a Logical Point of View*, pp. 20–46

[13] I trace the history in some detail in "Taking Naturalism Seriously," in *Proceedings of the 9th International Congress of Logic, Methodology and Philosophy of Science*, D. Prawitz, B. Skyrms, and D Westerstahl, eds. (Amsterdam: North Holland, forthcoming).

Some philosophers might be tempted to discount this behavior of actual scientists on the grounds that experimental confirmation is enough, but such a move is not open to the naturalist. If we remain true to our naturalistic principles, we must allow a distinction to be drawn between parts of a theory that are true and parts that are merely useful. We must even allow that the merely useful parts might in fact be indispensable, in the sense that no equally good theory of the same phenomena does without them. Granting all this, the indispensability of mathematics in well-confirmed scientific theories no longer serves to establish its truth.

But perhaps a closer look at particular theories will reveal that the actual role of the mathematics we care about always falls within the true elements rather than the merely useful elements; perhaps the indispensability arguments can be revived in this way. Alas, a glance at any freshman physics text will disappoint this notion. Its pages are littered with applications of mathematics that are expressly understood not to be literally true: e.g., the analysis of water waves by assuming the water to be infinitely deep or the treatment of matter as continuous in fluid dynamics or the representation of energy as a continuously varying quantity. Notice that this merely useful mathematics is still indispensable; without these (false) assumptions, the theory becomes unworkable.

It might be objected that these applications are peripheral, that they are understood against the background of more fundamental theories, and that it is in contrast with these that the applications mentioned above are "idealizations," "models," "approximations," or useful falsehoods. For example, general relativity is a fundamental theory, and when space-time is described as continuous therein, this is not explicitly regarded as less than literally true. So the argument goes.

But notice, when those pre-Einsteinians were skeptical of atomic theory, it was a fundamental theory in this sense; it was not proposed against another background as a convenient idealization or mere approximation. The skeptics were bothered, not by such peripheral simplifications, but by what they saw as the impossibility of directly testing the core hypotheses of atomic theory. But consider now the hypothesis that space-time is continuous. Has this been directly tested? As Quine himself points out, "no measurement could be too accurate to be accommodated by a rational number, but we admit the [irrationals] to simplify our computations and generalizations."[14] Similarly, space-time must be regarded as continuous so that the highly efficacious continuum mathematics can be applied to it. But the key question is this: Is that continuous character "experimentally verified" or merely useful? If it is merely useful, then the indispensability argument sketched earlier, the one relying on the role of continuum mathematics in science to support the Zermelo-Fraenkel axioms, or ZFC, cannot be considered conclusive.

I shall not try to answer this question here; to do so would require a more thorough study of the physics literature than I am capable of launching just now. But until such a study is undertaken, until the evidence for the literal continuity of space-time is critically examined,[15] I think the simple observations collected here are enough to raise a serious question about the efficacy of this particular indispensability argument.

IV. FOUNDATIONS OF SET THEORY

My second reservation about the indispensability arguments rests on somewhat less familiar grounds; to reach it, I must review a bit of set theory and (in the next section) a bit of physics.

It is well-known that the standard axioms of contemporary set theory, ZFC, are not enough to decide every naturally-arising set theoretical question.[16] The most famous independent statement is the Cantor's continuum hypothesis (CH), but there are others, some of them more down-to-earth than CH. For example, between the mid-seventeenth and the late nineteenth century, under pressure from both physical and mathematical problems, the notion of function became more and more general. Around the beginning of our century, various mathematicians undertook to bring order to the wild domain of discontinuous functions. It soon became clear that the complexity of functions could be understood in terms of the complexity of sets of real numbers—e.g., a function is continuous iff the inverse image of every open set is open—and this naturally led to a serious study of the properties of definable sets of reals. Among these, the Borel sets could easily be shown, for example, to be Lebesgue measurable. This result generalized to projections[17] of Borel sets (the analytic sets) and to the complements of these (the coanalytic sets). One more application of projection produces the sets we now call Σ^1_2, but the question of their Lebesgue measurability remained stubbornly unsolved. Sometime later, this question was shown to be independent of ZFC.[18]

---

[14] "Reply to Charles Parsons," p. 400.

[15] This issue of the continuity of space-time will take an unexpected turn in sect. V.

[16] Kurt Gödel's incompleteness theorem is enough to establish that there are set-theoretic statements that can neither be proved nor disproved from ZFC, but Gödel's later work on the inner model of constructible sets and Paul Cohen's forcing methods yield more: there are statements *which mathematicians have found it natural to ask* which are likewise independent.

In contrast with CH, this question concerns only a limited class of definable sets of reals, sets whose definitions have concrete geometric interpretations, and it involves the intuitive analytic notion of Lebesgue measurability rather than Cantor's bold new invention, the comparison of infinite cardinalities. In other words, it might be said that this independent question, unlike CH, arose in the straightforward pursuit of analysis-as-usual. And there are others of this type.

There is a serious foundational debate about the status of these statements. Despite their independence of ZFC, one might hold that there is nevertheless a fact of the matter, that the statements are nevertheless either true or false, and that it is the burden of further theorizing to determine which.[19] At the other extreme, another might insist that ZFC is all there is to set theory, that a statement independent of these axioms has no inherent truth value, that the study of extensions of ZFC that settle these questions one way or the other are all equally legitimate. For future reference, let me attach labels to crude versions of these positions: let *fact* be the bare assumption that there is a determinate answer to our question, and let the opposing view be *no-fact*.

Now let us pose this foundational question to the indispensability theorist, taking the simple version of indispensability first, for purposes of comparison.[20] Continuum mathematics, including everything from real-valued measurement to the higher calculus, is among the most widely applied branches of mathematics, and at least some of the many physical theories in which it is applied are extremely well-confirmed. Therefore, so the argument goes, we have good reason to believe in the entities of continuum mathematics, for example, the real numbers. In these applications—for example, in the theory of space (or space-time)—we also find quantification over sets of reals, (or equivalently, over regions of space (or space-time)), though particular instances are rarely as complex as Σ^1_2.[21] If we believe in the reals and in those sets of reals definable in our theory, then it seems we should accept the legitimacy of the question: Are Σ^1_2 sets Lebesgue measurable?[22] Thus the simple indispensability theorist should endorse *fact*.

---

[17] The projection of a subset of the plane is its shadow on one of the coordinate axes.

[18] I discuss this history in more detail in ch. 4 of *Realism in Mathematics* and in "Taking Naturalism Seriously."

[19] This was Gödel's view, and the one defended in *Realism in Mathematics*.

What follows from this for the practice of set theory? Should the set theorist, as Gödel suggests, seek an answer to this legitimate question? Given that our independent question seems, for now, to have no bearing on physical theory, and that it is not settled by the most generous of "simplificatory rounding outs" (i.e., ZFC), the simple indispensability theorist, who uses only the justificatory methods of physical science, has no means for answering it. Furthermore, the exclusive focus on the needs and methods of physical science hints at a lack of interest in any question without physical ramifications. If so, the simple theorist may disagree with *no-fact*, classifying our independent question as one with an unambiguous truth value, without going so far as charging future theorists with the task of answering it. Call this *weak fact*.

Fortunately, these subtle matters of interpretation are beside the point here, because we have already identified the modified indispensability argument as more promising than the simple. Like the simple theorist, the modified indispensability theorist embraces the ontology of continuum mathematics, on the basis of its successful applicability, and thus the legitimacy of our independent question, but she goes beyond the simple theorist by ratifying the set theorist's search for new axioms to answer the question. Call this *strong fact*. Although the evidence for or against these axiom candidates will derive, not from physical applications, but from considerations internal to mathematics, the modified theorist sees the past success of such mathematical methods as justifying their continued use.

---

[20] For now, I shall ignore the scientific practice objection and take fundamental scientific theory at face value.

[21] One theoretical proposal that involves nonmeasurable sets of reals is I Pitowsky, "Deterministic Model of Spin and Statistics," *Physical Review D*, XXVII, 10 (15 May 1983). 2316–26 (M. van Lambalgen called this paper to my attention.)

[22] This might be avoided if we took mathematical entities to be somehow "incomplete," an idea of Parsons's which Quine considers in passing See Quine's "Reply to Charles Parsons," p. 401, and references cited there

Either way, then, the indispensability theorist should adopt some version of *fact*. Notice, however, that this acceptance of the legitimacy of our independent question and (for the modified theorist) the legitimacy of its pursuit is not unconditional; it depends on the empirical facts of current science. The resulting mathematical beliefs are likewise a posteriori and fallible.

V. FOUNDATIONS OF PHYSICS

Here the fundamental theories are general relativity and quantum mechanics, and the central problem is to reconcile the two. Three of the fundamental forces—electromagnetic, weak, and strong—have yielded more or less workable quantum-field descriptions, but gravity remains intractable. The mathematics of quantum-field theories has an annoying habit of generating impossible (infinite) values for some physical magnitudes, but the problem has been overcome by a technical trick called "renormalization" in all cases but that of gravity. So the problem remains: How to characterize the gravitational force at quantum distances?

Physicists engage in fascinating speculations on the source of the difficulty: for example, that it arises from the attempt to combine the essentially passive space-time of quantum theory with the dynamic space-time of general relativity.[23] But most important for our purposes is the idea that the fault might lie in our conception of space-time as a mathematical continuum. For example, Richard Feynman[24] writes:

> I believe that the theory that space is continuous is wrong, because we get these infinities and other difficulties . . . (*ibid.*, p. 166).

And Chris Isham:

> . . . it is clear that quantum gravity, with its natural Planck length, raises the possibility that the continuum nature of spacetime may not hold below this length, and that a quite different model is needed (*op. cit.*, p. 72).

---

[23] See Chris Isham, "Quantum Gravity," in *The New Physics*, P. Davies, ed. (New York: Cambridge, 1989), pp. 70–93, esp. p. 70 I review some of the popular literature on this problem in "Taking Naturalism Seriously."

[24] *The Character of Physical Law* (Cambridge: MIT, 1967).

Here the suggestion is not (as in the previous section) that the continuity of space-time is a "mere idealization," but that it does not belong in our best theory at all!

All this, as I have indicated, is quite speculative; no one yet knows what a reasonable theory of quantum gravity might be like. But the very suggestion that space-time may not be continuous is enough to add previously unimagined poignancy to our earlier conclusion of a posteriority and fallibility. What seemed a rather small concession at the end of the last section—that the grounds for the indispensability theorist's adherence to *fact* could be overthrown by progress in physics—now looms as a real possibility.

Of course, the (potential) falsity of continuum mathematics in its application to space-time would be only part of the story; there are other successful scientific uses for the calculus and higher analysis. But, if science were to change so that all fundamental theories were thoroughly quantized, so that no continuum mathematics appeared there, if all the remaining applications of continuum mathematics were explicitly understood as "approximations" or "idealizations" or "models," then even the modified indispensability theorist would retreat to some version of *no-fact*. The case of quantum gravity should keep us from dismissing this possibility out of hand.

VI. THE MATHEMATICAL PRACTICE OBJECTION

For the modified indispensability theorist, the choice between *strong fact* and *no-fact* hinges on developments in physics (and perhaps the rest of science). We should now ask the impact of this choice: How would the pursuit of our independent question be affected by it? In other words, we want to know if the metaphysical distinction between *strong fact* and *no-fact* has methodological consequences.

If *no-fact* is correct, if there is no pre-existing fact to discover about the Lebesgue measurability of Σ^1_2 sets of reals, then what approach should the set theorist take? Many observers would hold that *no-fact* is the end of the story, that mathematicians are in the business of discovering truths about mathematical reality, and that, if there is no truth to be found, the mathematician should reject the question. From this point of view, all (relatively consistent) set theories extending ZFC are equally legitimate, there is no call to chose between them, and indeed, no grounds on which to do so apart from subjective aesthetic preferences. Once our question is shown to be independent, and developments in science undercut the claim to inherent truth value, there is nothing more of serious import to be said about the Lebesgue measurability of Σ^1_2 sets. Call this *end of the story no-fact*.

This position makes nonsense of the contemporary search for new set-theoretic axioms to settle independent questions like ours. Indeed, in our case, there are two competing candidates: $V = L$ (Gödel's "axiom of constructibility") and MC (the existence of a measurable cardinal). If $V = L$, then there is a non-Lebesgue measurable Σ^1_2 set; if MC, then all Σ^1_2 sets are Lebesgue measurable. Set theorists offer arguments for and against these axiom candidates,[25] and in this debate, MC is strongly favored over $V = L$.[26] If we are not to reject this activity as inconsequential mutterings—an especially unappealing move, given that the original axioms of ZFC are supported by arguments of a similar flavor—we must instead reject *end of the story no-fact*.

But there is another version of *no-fact*. Even if there is no pre-existing fact of the matter to be discovered, the process of extending the axioms of set theory might well be governed by nonarbitrary principles. This idea turns up, not only in the study of set theory, but when ontological decisions are made in other branches of mathematics as well. For example, Kenneth Manders[27] describes the theoretical norms at work in the expansion of the domain of numbers to include the imaginary or complex numbers, and Mark Wilson[28] uncovers the rationale behind the move from affine to projective geometry. In such cases, despite lip service to the notion that any consistent system is as good as any other, mathematicians actually insist that a given mathematical phenomenon is correctly viewed in a certain (ontological) setting, that another setting is incorrect.

One need not assume *fact* to endorse these practices. Even if there is no fact of the matter, no pre-existing truth about the existence or nonexistence of complex numbers or geometric points at infinity or nonconstructible sets, the pursuit of these mathematical topics might be constrained by mathematical canons of "correctness." For our case, one might hold that there is no fact of the matter about the Lebesgue measurability of Σ^1_2 sets, but that there are still good mathematical reasons to prefer extending ZFC in one way rather than another, and perhaps, good mathematical reasons to adopt a theory that decides our question one way rather than another. From this point of view, *no-fact* is just the beginning of the story; it opens the door on the fascinating study of purely mathematical canons of correctness. Call this *beginning of the story no-fact.*

---

[25] See my "Believing the Axioms I–II," *The Journal of Symbolic Logic*, LIII, 2 (June 1988): 481–511, and 3 (September 1988): 736–64.

[26] I discuss part of the case against $V = L$ in "Does V equal L?" in *The Journal of Symbolic Logic* (forthcoming).

[27] "Domain Extension and the Philosophy of Mathematics," this JOURNAL, LXXXVI, 10 (October 1989): 553–62.

[28] "Frege: The Royal Road from Geometry," in *Noûs* (forthcoming)

The question before us is this: What are the methodological consequences of the choice between *strong fact* and *no-fact?* If *strong fact* is correct, the set theorist in search of a complete theory of her subject matter should seek out additional true axioms to settle the Lebesgue measurability of Σ^1_2 sets (and so on). If *end of the story no-fact* is correct, the set theorist left with any interest in the matter should feel free to adopt any (relatively consistent) extension of ZFC she chooses, or even to move back and forth between several mutually contradictory such extensions at will. And finally, if *beginning of the story no-fact* is correct, the set theorist should use appropriate canons of mathematical correctness to extend ZFC and to decide the question.[29]

Obviously, the method prescribed by *strong fact* differs from that prescribed by *end of the story no-fact;* that much is easy. But what about *strong fact* and *beginning of the story no-fact?* Does the pursuit of truth differ from the pursuit of mathematical correctness? In fact, I think it does. Consider, for example, a simple argument that $V = L$ should be rejected because it is restrictive. A supporter of this argument owes us an explanation of why restrictive theories are bad. A *beginning of the story no-fact*-er might say, "because the point of set theory is to realize as many isomorphism types as possible, and set theory with MC is richer in this way."[30] A *strong fact*-er might agree that the world of MC has desirable properties, while insisting that desirability (notoriously!) is no guarantee of truth.[31] Faced with the *beginning of the story no-fact*-er's argument, a *strong fact*-er would reply, "Yes, MC is nice in the way you indicate, but if *V does* equal L, L *does* contain all the isomorphism types possible. What's needed is an argument that $V = L$ is *false*." So *strong fact* will differ methodologically from *beginning of the story no-fact* as well as *end of the story no-fact*.

---

[29] The mathematical canons invoked in *beginning of the story no-fact* could ultimately recommend that several different set theories be accorded equal status. The methodology at work would still be different from that of *end of the story no-fact,* and the range of theories endorsed would almost certainly be narrower.

[30] Spelling out this line of thought precisely is no simple exercise, but I shall leave that problem aside here. The point is just that the *beginning of the story no-fact*-er appeals to some attractive feature of set theory with MC.

[31] Philip Kitcher touches on this point in his reply to Manders, "Innovation and Understanding in Mathematics," this JOURNAL, LXXXVI, 10 (October 1989): 563–4, when he writes, p. 564: "Suppose this is a way in which mathematical knowledge can grow. What kinds of views of mathematical reality and mathematical progress are open to us? Can we assume that invoking entities that satisfy constraints we favor is a legitimate strategy of recognizing hitherto neglected objects that exist independently of us? From a realist perspective, the method of postulating what we want has (in Bertrand Russell's famous phrase) 'the advantages of theft over honest toil.' If that method is, as Richard Dedekind supposed, part of the honest trade of mathematics, is something wrong with the realist perspective?"

We have reached this point: a methodological decision in set theory—namely, that between the methodologies proper to *strong fact* and to *beginning of the story no-fact*—hinges on developments in physics. If this is correct, set theorists should be eagerly awaiting the outcome of debate over quantum gravity, preparing to tailor the practice of set theory to the nature of the resulting applications of continuum mathematics. But this is not the case; set theorists do not regularly keep an eye on developments in fundamental physics. Furthermore, I doubt that the set-theoretic investigation of independent questions would be much affected even if quantum gravity did end up requiring a new and different account of space-time; set theorists would still want to settle open questions about the mathematical continuum. Finally, despite the current assumed indispensability of continuum mathematics, I suspect that the actual approach to the Lebesgue measurability of Σ^1_2 sets, to $V = L$ versus MC, is more like that prescribed by *beginning of the story no-fact* than that prescribed by *strong fact*,[32] and I see no mathematical reason to criticize this practice. In short, legitimate choice of method in the foundations of set theory does not seem to depend on physical facts in the way indispensability theory requires.

VII CONCLUSION

I have raised two doubts about indispensability theory, even modified indispensability theory, as an account of mathematics as practiced. The first, the scientific practice objection, notes that indispensability for scientific theorizing does not always imply truth and calls for a careful assessment of the extent to which even fundamental mathematized science is "idealized" (i.e., literally false). The second, the mathematical practice objection, suggests that indispensability theory cannot account for mathematics as it is actually done. If these objections can be sustained, we must conclude that the indispensability arguments do not provide a satisfactory approach to the ontology or the epistemology of mathematics. Given the prominence of indispensability considerations in current discussions, this would amount to a significant reorientation in contemporary philosophy of mathematics.

PENELOPE MADDY

University of California/Irvine

---

[32] I hope to argue this in some detail elsewhere.

THE SKELETON IN FREGE'S CUPBOARD: THE STANDARD VERSUS NONSTANDARD DISTINCTION

Anyone who uses higher-order logic faces a momentous choice concerning the interpretation of higher-order quantifiers. Let us consider as an example a second-order quantifier that involves a one-place class or predicate variable, say, X. Its values can be taken to be either classes of individuals or properties (concepts) of individuals. In either case, the same dilemma confronts a higher-order logician, even if its horns look different on the two interpretations.

If the values of X are thought of as being classes, then the crucial question is whether the range of a quantifier containing X is the entire power set $P(do(M))$ of the relevant domain $do(M)$ of individuals, or only some designated subset of $P(do(M))$. In other words, the question is whether the values of X are arbitrary extensionally possible classes or whether only some such classes are accepted as values of X. The former alternative results in what is usually called the *standard* interpretation of higher-order logic, the latter in a *nonstandard* interpretation.

A nonstandard interpretation can be of many different kinds. The relevant range of X is sometimes thought of as a part of the specification of the model M in which a higher-order formula is being interpreted, sometimes as a restraint on the interpretation. Often, the range of X and the ranges of other higher-order variables cannot be selected completely arbitrary, but are subject to certain closure conditions. For instance, the totality of these different ranges is assumed to be closed with respect to Boolean operations and to projective ones; in other words, the usual formation rules are assumed to preserve interpretability.[1]

The same distinction has to be made for variables of each logical type. An especially important case in point is the interpretation of function variables, as witnessed by the fact that a specific nonstandard interpretation is sometimes imposed on them on purpose. The

---

[1] Leon Henkin, "Completeness in the Theory of Types," *The Journal of Symbolic Logic*, XV (1950): 81–91. For an important correction to Henkin's paper, see Peter B. Andrews, "General Models and Extensionality," *The Journal of Symbolic Logic*, XXXVII (1972): 395–7.

0022-362X/92/8906/290–315

best known essay in this direction is probably Kurt Gödel's[2] functional interpretation of first-order logic and arithmetic. There the values of function variables are restricted to recursive functions of the appropriate type. Notwithstanding the variety of different types of entities and their intricate interrelations, the apparently special case of one single second-order quantifier with a one-place predicate or class variable is fully representative of the entire theory of finite types. For there are results that show that, if the standard interpretation is granted to one single second-order one-place quantifier, the entire higher-order logic (theory of finite types) with the standard interpretation can be reconstructed by its means.[3] Of course, by the same token the interpretation of a single-function variable determines in a sense the interpretation of the entire higher-order logic, for classes can always be handled by means of their characteristic functions.

Philosophers have not paid adequate attention to the standard versus nonstandard distinction. The main reason is undoubtedly their slowness to appreciate the importance of the model-theoretical viewpoint in logical theory.[4] Be this as it may, the contrast was spelled out explicitly by Leon Henkin in 1950. The fact that he considered only one nonstandard interpretation does not detract from his achievement.

I. CONCEPTS VERSUS THEIR EXTENSIONS

The interpretational dilemma cannot be avoided by switching to variables that range over properties or relations in contradistinction to their extensions. Then the question whether the standard interpretation is assumed becomes the question as to whether for each class (potential extension), *C*, there is a property that has *C* as its extension or, in Gottlob Frege's term, as its course-of-values.

This distinction is extended as a matter of course to variables of other higher-order types. As a bargain, we are put into the position of being able to distinguish the main problem addressed in this paper from two others. In this paper, we are asking whether Frege assumed that for each potential extension there is a concept that has this extension as its value range. One can also ask the mirror-image question: Is there an extension for each concept? It turns out to be extremely important to separate the two questions from each other, both in discussing Frege and in discussing other logicians and mathematicians (cf. below).

---

[2] "On a Hitherto Unexploited Extension of the Finitary Standpoint," *Journal of Philosophical Logic*, IX (1980): 133–42

[3] For such results, see Hintikka, "Reductions in the Theory of Types," *Acta Philosophica Fennica*, V (1955): 59–115

[4] This is documented in Hintikka, "On the Development of the Model-Theoretic Viewpoint in Logical Theory," *Synthese*, LXXVII (1988): 1–36

We can also formulate a third question: Are two concepts with the same extension identical? This question is often combined with the first two and formulated as the query: Is logic extensional? It turns out, however, that often the philosophers who ask this question really have in mind only one of the three questions. It is also important to realize that an "extensionalist" answer to the question concerning the identity conditions of functions (including concepts), which is Frege's answer, by no means prejudices his answer to either of the other two questions.

II THE STANDARD VERSUS NONSTANDARD DISTINCTION AND THE EXPRESSIBILITY OF CONCEPTS IN LANGUAGE

Sometimes, a nonstandard interpretation is guided by the idea that only such properties, relations, and functions can be assumed to exist as can be defined or otherwise captured by a suitable expression of one's language. In the case of theories with infinite models, this leads inevitably to a nonstandard interpretation, for there can be only a countable number of such definitions or characterizations available for this purpose. Hence they cannot capture all the subsets of *do*(*M*), for there is an uncountable number of them.

This particular nonstandard interpretation was the only one considered by Henkin. More generally, it seems to be the only one that is considered in most of the literature where the Henkin distinction is relied on or discussed. It is nevertheless important to realize that other nonstandard interpretations are possible. What is crucial is to realize that the nonstandard versus standard distinction is not the same as the distinction between, on the one hand, variables each of whose values always has a representative in language which can serve as a substitution value for the variable and, on the other, variables whose values can exist independently of their linguistic representations. The latter issue can be called the problem of nonlinguistic existence (of values of higher-order variables). Historically, it may have been the source of the standard versus nonstandard distinction, but, systematically, it is independent of this distinction. In general, representability in a language on the basis of a theory depends on the language and on the theory, while the choice between the standard and a nonstandard interpretation faces us in using any higher-order language.

A confusion between the different questions we have just distinguished is not merely an abstract possibility. The fallacy of not separating them has been committed even by knowledgeable philosophers. For instance, in an article in *Erkenntnis,* Rudolf Carnap[5] criticizes Frank Ramsey's reliance on the standard interpretation and writes:

> I think we should not let ourselves be seduced by it into accepting Ramsey's basic premise, viz., that the totalities of properties already exist before their characterization by definition. . .
> I think we ought to hold fast to Frege's dictum that in mathematics only that may be taken to exist whose existence has been proved (and he meant proved in finitely many steps) (*ibid.,* p. 102).

Here Carnap is clearly misinterpreting Ramsey's position. In so far as we are dealing with what Ramsey called predicates and relations "in the ordinary sense," Ramsey apparently accepts the idea that there has to be an expression for them. In this sense, Ramsey in fact discusses how functions are constructed by constructing expressions for them. Ramsey's point is that our entire logic has to be reinterpreted so as to deal, "not with relations in ordinary sense, but with possible correlations, or 'relations in extension' as Mr. Russell calls them," and likewise for entities of other logical types. In fact, Carnap seems to misinterpret Frege's position, too, or at least he gives a wrong description of it, as we shall see.

A terminological point is in order here. By calling one of the two contrasting interpretations "standard," we are not passing any judgment as to which view ought to be adopted or which one is historically the usual one. We are merely following the terminology of Henkin, who was the first logician to articulate fully the standard versus nonstandard distinction.

III. THE STANDARD VERSUS NONSTANDARD DISTINCTION AND SET THEORY

On the basis of what has been said, it might seem that the standard interpretation has set theory as its main fortress. In the first section of this paper, we explained what the standard interpretation is by means of the familiar set-theoretical notion of the power set. It is undoubtly true that Georg Cantor intuitively assumed some sort of standard interpretation in his thinking. It can even be argued, as Michael Hallett[6] has aptly done, that the actual generated principles for transfinite ordinals go "most of the way toward acceptance of the existence of *arbitrary* sub-domains" and therefore toward the standard interpretation. As Hallett writes:

---

[5] "Die logizistische Grundlegung der Mathematik," *Erkenntnis,* II (1931): 91–105.

[6] *Cantorian Set Theory and Limitation of Size* (New York: Oxford, 1990).

> Cantor asserts that the two principles of generation [he had formulated] 'give us the ability to break through every barrier in the formation of real, whole numbers' (the italics are Cantor's). This certainly suggests a belief that these methods take the ordinal number sequence arbitrarily far (*ibid.*, p. 58).

But the real question here is whether later set theorists were able to implement Cantor's intuitive ideas so as to turn what Hallett guardedly refers to as "most of the way" into all the way to Cantor's implicitly "standard" way of thinking about sets and their existence. The right answer to this question is a resounding *no*.

One way of looking at contemporary axiomatic set theory is to view it as an attempt to deal with such higher-order entities as sets on the first-order level. What inevitably gets lost in such an attempt is precisely the possibility of a standard interpretation of variables ranging over higher-order entities. This problem manifests itself, among other things, in the form of the Skolem paradox. Even when you seem to be able to prove the existence of all sorts of uncountable sets in your axiom system, it turns out that there are, according to the downward Löwenheim-Skolem theorem, models of the same axiom system in which those very sets are countable. Indeed, the very paradox about the Löwenheim-Skolem theorem lies in the fact that what in the system is called countability or uncountability is that only on the standard interpretation of set variables understood as higher-order ones.

It does not automatically help, either, to move to a higher-order level: higher-order logics can typically be treated as if they were many-sorted first-order logics *as long as the standard interpretation is not assumed*. Thus, in a sense it is the standard versus nonstandard interpretation that constitutes the real watershed between first-order logic and higher-order logics. At least this distinction is much more consequential than a mere distinction between types (order levels). Since Frege's logic was a higher-order one, we are already beginning to see that the standard versus nonstandard distinction puts his entire project into a sharp new light.

IV. THE IMPORTANCE OF THE DISTINCTION

We called the standard versus nonstandard distinction momentous. This was in reality an understatement. Each horn of the dilemma turns out to be very sharp, in that the prospects of higher-order logic and of its applications depend essentially on one's choice between the two kinds of interpretations, in more than one way.

For one important thing, second-order logic with a standard interpretation is easily seen to be semantically incomplete. In other words, the set of valid formulas is not recursively enumerable on this interpretation. There cannot exist a finite (or recursively enumerable) set of logical axioms and inference rules that would generate all and only valid formulas as theorems. In contrast, Henkin showed that, on one especially natural nonstandard interpretation, higher-order logic does after all admit of a semantically complete axiomatization (*op. cit.*, p. 81). Many other kinds of nonstandard interpretation likewise yield axiomatizable logics.[7] This does not mean that we should prefer a nonstandard interpretation. For only the standard interpretation of second-order logic enables us to use it for the most important purposes it can serve in the foundations of mathematics. If we assume the standard interpretation, we can easily formulate descriptively complete and indeed categorical axiomatizations for such crucial mathematical theories as number theory and the theory of real numbers.[8] Likewise, the principle of mathematical induction is easily formulated in a second-order logic with standard interpretation, but cannot be formulated in an axiomatizable nonstandard higher-order logic.[9] Thus, for most serious foundational purposes we have to assume the standard interpretation of second-order logic.[10]

V. THE ROLE OF THE DISTINCTION IN FOUNDATIONAL DISCUSSIONS

Even though the standard versus nonstandard distinction has not attracted much attention from philosophers, it has played an important, albeit only partially articulated, role in the actual foundational discussions. For instance, the standard versus the nonstandard contrast is the most important determinant of a mathematician's attitude to the axiom of choice. On the standard interpretation, the axiom of choice is trivially acceptable, whereas for an advocate of a nonstandard interpretation, there is prima facie little reason to believe in the axiom of choice. This important background of the career of the axiom of choice is almost universally neglected.[11]

---

[7] See also Hintikka, "Standard vs. Nonstandard Logic· Higher Order, Modal and First-Order Logics," in *Modern Logic: A Survey*, E. Agazzi, ed. (Boston: Reidel, 1980), pp. 283–96.

[8] Henkin, p. 89. See also Steward Shapiro, *Foundations without Foundationalism* (New York: Oxford, 1991).

[9] For this point, see Shapiro, and also Thoralf Skolem, "Über einige Grundlagenfragen der Mathematik," *Skrifter utgitt av det Norske Videnskaps-Akademi*, I, 4 (1929): 1–49.

[10] The whole business of axiomatizability is nevertheless extremely subtle. Perhaps the most interesting nonstandard interpretation of first-order logic is obtained by restricting Skolem functions to recursive ones, somewhat as in Gödel's functional interpretation, but without any unusual interpretation of negation and conditionals. Then the resulting logic is nonaxiomatizable, as was first pointed out by Robert Vaught in "Sentences True in All Constructible Models," *The Journal of Symbolic Logic*, XXV (1960): 39–53. The Peano axiomatization of elementary arithmetic is then complete, however, as was first pointed out by Tennenbaum. For this later point, see Richard Kaye, *Models of Peano Arithmetic* (New York. Oxford, 1991), esp. pp. 153–7.

Likewise, the standard versus nonstandard distinction prejudices one's judgment of Bertrand Russell's axiom of reducibility, which roughly says that for each impredicatively defined set there exists an equivalent predicatively definable class. For a believer in the standard interpretation, a class is there, determined by its members, no matter how it is defined in some one language. Hence, for the theorist of the standard sort, the "axiom of reducibility" is trivially true, indeed redundant, as was forcefully pointed out by Ramsey.[12] In general, Ramsey's revision of Russell's ramified theory of types is to all intents and purposes a self-conscious step from a nonstandard to a standard higher-order logic.

VI. STANDARD INTERPRETATION AND THE IDEA OF ARBITRARY FUNCTION

We can say that there were two long-range developments of mathematicians' thinking about the foundations of their own field which led to the idea of standard interpretation:

(i) The development of the idea of an arbitrary function.
(ii) The development of the idea of arbitrary set.

Thinking about (i) and (ii) was conditioned not so much by philosophical questions as by the demands of mathematical theorizing itself. Let us turn first to the notion of arbitrary function.

The development of the idea of an arbitrary function is one of the most important conceptual developments in mathematics. This development has not always been fully appreciated. For instance, the concept of arbitrary function is often credited to P. G. L. Dirichlet, even though in reality it was formulated as early as 1755 by L. Euler. In his *Institutiones calculi differentialis,* Euler wrote:

> If some quantities so depend on other quantities that is the latter are changed the former undergo change, then the former quantities are called functions of the latter. This denomination is of the broadest nature and comprises every method by means of which one quantity

---

[11] This is the case even with Gregory Moore's important book, *Zermelo's Axiom of Choice: Its Origins, Development and Influence* (New York: Springer, 1982).

[12] "The Foundations of Mathematics," *Proceedings of the London Mathematical Society,* II, 25, (1925): 338–84.

> could be determined by others. If, therefore, x denotes a variable quantity, then all quantities which depend upon x in any way or are determined by it are called functions of it.[13]

Similar quite general characterizations of a function are found in other mathematicians, for instance in N. I. Lobachevsky. According to I. H. Anellis,[14] Lobachevsky thought that "general thought demands that a function [f] of *x* be called a number which is given for each *x* and which changes . . . together with *x*." The case of Lobachevsky is interesting because the development of non-Euclidean geometries was one of the factors that prompted mathematicians to consider a wider class of structures than before: for instance, because not all non-Euclidean geometries can be modeled within the Euclidian one.

In Euler's time, the concept of arbitrary function had little use in state-of-the-art mathematical research. Actual work both in the foundations of mathematics and at the cutting edge of the subject gradually raised the consciousness of most mathematicians in the course of the nineteenth century. The innocent looking Eulerian definition of an arbitrary function turned out to admit of surprisingly unruly denizens of the world of mathematics. For instance, many of the results that had been taken more or less for granted turned out to hold only for functions with specifiable properties. For instance, it turned out that even a continuous function need not be differentiable. This development enhanced, on the other hand, the interest of such results as could be proved for arbitrary functions or at least for a large class of functions. Important examples are provided by functions that can be represented by means of Fourier series and by functions that are integrable, on one of the various extended definitions of integration.

As a consequence of such developments, a lively interest arose in the concept of function and in the question as to how wide a class of functions (possibly including all arbitrary functions) mathematicians should be concerned with. The acceptance or rejection of the concept of arbitrary function was, for such reasons, an important watershed in the foundational discussions in the nineteenth century.

---

[13] This passage from *Institutiones calculi differentialis* is translated by A. P Yushevich: "The Concept of Function up to the Middle of the 19th," *Arch. Hist. Exact. Sci.*, XVI (1970), p. 70. For additional information on the development of the idea of an arbitrary function, see also Umberto Bottazzini, *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, (New York: Springer, 1986).

[14] *A History of Mathematical Logic in Russia and the Soviet Union*, unpublished.

The tension generated by this question is in evidence in several leading mathematicians. For instance, K. T. Weierstrass considered the idea of arbitrary function, but rejected it as being so vague and so general as not to enable us to find any interesting things to say about it as such. The other side of the Weierstrassian idea of function, however, was his desire to consider as large a class of functions as possible, to find "the largest class of functions for which one can give an analytic representation and which can most fully satisfy the needs of analysis."[15]

L. Kronecker, in contrast, rejected clearly the idea of an arbitrary function. As Harold M. Edwards[16] puts it, "what is missing, and what is excluded by his [Kronecker's] principles is integrals of 'arbitrary' functions or sums of 'arbitrary' infinite series" (*ibid.*, p. 74).

All these developments are highly relevant here, for the conception of the standard interpretation of higher-order quantifiers (and variables) is, to all purposes, equivalent with the notion of an arbitrary function or the notion of an arbitrary set, as was indicated in section I above. For the same reason it is not anachronistic to relate foundational discussions of the late nineteenth century to the standard versus nonstandard distinction, even though it entered into actual discussion disguised by a different terminology.

VII. PRIMACY OF CONCEPTS OVER THEIR EXTENSIONS IN FREGE

After all these preparations, we can turn to the question whether Frege assumed the standard interpretation of higher-order quantifiers or a nonstandard one. This question is especially poignant when applied to Frege, for his logic contains explicitly formulated higher-order quantifiers. We shall show that Frege lacked both the idea of arbitrary function and the idea of arbitrary set, and hence in effect opted for a nonstandard interpretation of higher-order quantifiers. Let us deal first with the latter question.

For Frege, classes as classes come into play only through being *Wertverläufe* of functions (in Frege's wide sense of the term 'function'). He distinguished very clearly a concept, denoted by a concept word '$\Phi(\xi)$' from its extension, denoted by a course-of-values-name

---

[15] See P Dugac, "Eléments d'analyse de Karl Weierstrass," *Archive for History of Exact Sciences*, x (1973): 41–176.

[16] "Kronecker's view on the Foundations of Mathematics," in *The History of Modern Mathematics, Vol. I: Ideas and Their Reception*, David E. Rowe and John McCleary, eds. (San Diego: Academic, 1989)

'$\epsilon\Phi(\epsilon)$'.[17] As is well-known, one basic difference in Frege between a concept and the extension corresponding to it is that the latter is an object whereas the former is not. A concept is a function whose value is always a truth value. The extension of the concept, or alternatively the course of values of the function which the concept is, is an object that is determined by the arguments and the corresponding values of the function for these arguments. It is important to keep in mind that both concepts and courses of values are extensional (*ibid.*, pp. xxxix, 36).

The relation between a concept and its course of values is, contrary to what one might think, logical and not empirical. That is, the extension of a concept can only be apprehended by our logical faculties starting out from the concept.[18] Accordingly, a course of values does not exist independently, but only tied to a concept (function) which is logically more primary than it:

> . . . In fact I do hold that the concept is logically prior to its extension, and I regard as futile the attempt to base the extension of a concept or a class not on the concept but on individual things.[19]

Consequently, we have no means of recognizing a course of values except through a concept:

> We have only a means of always recognizing a course-of-values if it is designated by a name like "$\epsilon\Phi(\epsilon)$", by which it is always recognizable as a course-of-values. But we can neither decide, so far, whether an object is a course-of-values that is not given us as such, and to what function it may correspond, nor decide in general whether a given course-of-values has a given property unless we know that this property is connected with a property of the corresponding function.[20]

It has been argued by Tyler Burge[21] that, in framing his conception of class, Frege was the member of two traditions. The first was the intensional tradition originated by Leibniz and continued by J. S. Mill, G. Boole, and E. Schröder. Roughly, the main conception embodied in this tradition was that intension, in the form of properties, characteristic marks, etc., determine extension. The second tradition was the extensional one represented, among others, by Cantor, Richard Dedekind, and others. The main feature of this approach was that sets and numbers are not defined by appeal to properties, concepts, etc., as in the intensional tradition, but by "abstraction" (cf. section x below).

---

[17] Frege, *The Basic Laws of Arithmetic: Exposition of the System*, Montgomery Furth, trans. (Berkeley: California UP, 1967), p. 36.

[18] Frege, *Posthumous Writings*, H. Hermes, ed , P. Long and R. White, trans. (Cambridge: Blackwell, 1979), p. 181

[19] Frege, "A Critical Elucidation of Some Points in E. Schroeder's *Algebra der Logik*," in *Translations from the Philosophical Writings of Gottlob Frege*, P Geatch and M. Black, eds. (New York: Oxford, 1952), pp. 86–106, esp p. 106.

[20] Frege, *The Basic Laws of Arithmetic*, p. 46.

[21] "Frege on Extensions of Concepts, From 1884 to 1903," *Philosophical Review*, LXIII (1984): 1–34.

In his definition of a class, Frege adopted the former conception and he continuously criticized the extensionalists' idea that a class is formed by grouping together elements. It is important to realize that this antiextensionalism is compatible with Frege's view that the identity conditions for concepts are extensional. Frege's "intensionalism" (i.e., his view that concepts are logically prior to their extensions) can be seen in different ways. For instance, in his letter of July 24, 1902, Russell wrote Frege that he had difficulties "to see what a class really is if it does not really consist of objects but is nevertheless supposed to be the same for two concepts with the same extension."[22]

Frege's answer is contained in his letter to Russell of July, 28, 1902. It is worth spelling out in some detail. He distinguished wholes (systems) from classes, using the following examples, among others:

(1) The Romans conquered Gaul.
(2) The class of prime numbers comprises infinitely many objects.

In the first case, according to Frege, the Romans form a whole, that is, a physical body held together by relations that are essential to it, such as institutions, customs, laws. It is characteristic of wholes, Frege writes, that the parts of any part of a whole are again parts of the whole. Thus, the parts of the regiments are the battalions, but also the companies and the soldiers.

This is not so for classes, Frege avers. A class is determined by its members. The only members of the class of prime numbers are the prime numbers. The members of a class of regiments include only the regiments, not any individual soldiers. A class is a logical object, not a physical object like a whole (system). Asked how we apprehend classes, Frege gave the following unambiguous answer:

> We apprehend them as extensions of concepts, or more generally, as ranges of values of functions (*ibid.*, p. 141).

---

[22] Frege, *Philosophical and Mathematical Correspondence*, Gottfried Gabriel et al., eds. (Cambridge Blackwell, 1980), p. 139.

(Here and below, "ranges of values of functions" is synonymous with "courses of values of functions.")

This is not supposed to be interpreted as the priority of concepts only in explaining or understanding what classes are, but as an ontological principle as well:

> . . . for me the extension of a concept or a class is only a special case of a range of values (*ibid.*, p. 142).

Thus, a class which is not the extension of some concept is a contradiction in terms within the Fregean system, since every class has to be introduced via a concept. A class cannot be given through its elements, only through a rule that determines elementhood.[23]

Now, there were at least two different reasons for Frege to reject the individuation of classes through their members. The first one is connected with the definition of the empty class:

> Of course, one must not then regard a class as constituted by the objects (individual, entities), that belong to it; for removing the objects one would then also be removing the class constituted by them. Instead, one must regard the class as constituted by the characteristic marks, i.e., the properties which an object must have if it is to belong to it. It can then happen that these properties contradict one another, or that there occurs no object that combines them in itself. The class is then empty but without being logically objectionable for that reason.[24]

The second reason has to do with the individuation of infinite classes. Frege thought that an infinite class cannot be given solely by its members because of the finiteness of the human intellect. Instead, such a class can be justified only as derived from a concept involved in thought.[25]

VIII. FREGE LACKED THE NOTION OF ARBITRARY SET

The intensionalists' and, in particular, Frege's conception of set explained above was criticized by the so-called extensionalists, that is, set theorists like Cantor and Dedekind who shared with each other a different concept of set. Their main criticism against Frege was that he allowed an extension for every concept, which led to Russell's paradox. For instance, in his review of Frege (1884), Cantor[26] criticized Frege for defining classes as extensions of concepts. He pointed out that, in general, "the 'extension of a concept' is something quantitatively completely undetermined" and for this reason not all extensions of concepts are sets" (*ibid.*, pp. 440–1).

---

[23] See Frege, *Posthumous Writings*, p. 34; and *The Foundations of Arithmetic: A Logico-Mathematical Inquiry into the Concept of a Number*, J. L. Austin, trans (Cambridge: Blackwell, 1960).

[24] See the undated letter to Peano, in Frege, *Correspondence*, p. 109

[25] *Posthumous Writings*, p. 34; and *Foundations*, p. 98.

[26] "Rezension der Schrift von G. Frege *Die Grundlagen der Arithmetik*," in *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, E. Zermelo, ed. (Berlin: Springer, 1932)

In the same spirit, E. Zermelo[27] wrote:

> . . . it follows further as already appears in a much simpler way from Russell's antinomy, to be sure that it is not permissible to treat the extension of every arbitrary notion as a set and therefore that the customary definition of a set is too wide (*ibid.*, p. 107).

Even if Frege's definition of a set was too wide in one direction, in another direction it was too narrow. The inseparable Fregean linkage between sets and their defining properties prevented him from generating new sets by operations acting solely on the members of the already existing sets. This fact is correctly emphasized by Hao Wang[28] when he writes:

> He [Frege] identifies sets with their extensions and treats them as individuals (on the same level with individuals). This suggests immediately the idea of a type hierarchy of extensions, since extensions of predicates seem to be more closely related to predicates than to individuals . . . . But such a conception has little that is positive to say about how sets are to be generated (*ibid.*, p. 210).

Frege was thus unable to operate with arbitrary extensions as Cantor and Dedekind did. What is more, he tied extensions so closely to concepts that he could not reach the modern (generative) concept of set which or more or less completely implemented by the axiomatized set theories of Zermelo, J. von Neumann, and others. Although in these theories strong limitations on the formation of sets still exist (as compared with Cantor's system), still one has some liberty, starting with a given set, to form new sets. Thus, one basic ingredient of the standard interpretation, i.e., the notion of arbitrary set, is lacking from Frege's system.

One might still claim that, although Frege could not form arbitrary sets merely by applying certain operations on the members of an already existing set, there is nothing in Frege's theory which prevents the arbitrary sets to exist as value ranges of some suitable concepts. Hence, for Frege, the question whether the standard interpretation is correct becomes simply and solely the question whether there is a concept (function) for each potential extension, that is, whether for each class (class-in-extension) there is a concept whose course of values this class is. Thus, the answer to the question "Which of the two interpretations Frege was committed to?" depends crucially on Frege's conception of function.

---

[27] "Neuer Beweis für die Möglichkeit einer Wohlordnung," *Mathematische Annalen*, LXV (1908): 107–28.

[28] *From Mathematics to Philosophy* (New York: Routledge, 1974).

IX. THE OBJECTIVE EXISTENCE OF FUNCTIONS DOES NOT IMPLY STANDARDNESS

Michael Dummett[29] claims unambigously, not only that Frege operated with higher-order quantification, but that for Frege the range of a function quantifier included all "arbitrary functions":

> For Frege had not the slightest qualm about the legitimacy or intelligibility of higher-level quantification: he used it from the first, in *Begriffsschrift,* freely and without apology, and did not even see first-order logic as constituting a fragment having any special significance (*ibid.,* p. 223).

> And it is true enough, in a sense, that, once we know what objects there are, then we also know what functions there are, at least, so long as we are prepared, as Frege was, to admit all 'arbitrary functions' defined over all objects (*ibid.,* p. 177).

> That Frege can give no consistent justification for interpreting second-level quantification as varying over the full impredicative totalities of first-level concepts, relations and functions is, of course, no ground for denying him the right to employ second-level quantification at all (*ibid.,* p. 541).

Dummett does not provide any references to Frege's texts in support of his claims. It is nevertheless not difficult to see what he has in mind. On page 177, Dummett is expounding Frege's idea of the objective existence of functions, his "conception of the function as belonging to the 'realm of reference'." He points out that "for Frege the only possible route to grasp what a function is is via an understanding of the working of functional expressions." But this does not mean, Dummett in effect avers, that functions according to Frege are merely counterparts to functional expressions, that they are not inhabitants of the real universe. Dummett is undoubtly correct here. But what he is doing is merely a way of asserting the objective, nonlinguistic existence of functions independently of language. They are, in Dummett's words, mappings of objects on objects. Hence Dummett's real point is that Frege maintains the nonlinguistic existence of functions. Nothing follows from this point concerning Frege's alleged recognition of arbitrary function, i.e., of any possible correlation of objects and arguments. Dummett's move from the mode of existence of functions as objective correlations of objects to other objects to the claim that there was such a correlation or mapping for each possible pairing of argument values and function values is a *non sequitur*. He is apparently assuming that there is only one possible nonstandard interpretation, the one Henkin later considered. For only on this assumption could Dummett infer from Frege's rejection of this particular nonstandard interpretation that Frege accepted the standard one. At best, Dummett manages to refute Carnap's view (see note 5 above) that, according to Frege, only provably existing functions really exist.

---

[29] *Frege's Philosophy of Language* (London. Duckworth, 1973).

The other quotes from Frege which Dummett uses are not any more convincing. On the contrary, on page 541, Dummett even admits that Frege's "theory of reference cannot of itself, provide any temptation to violate the vicious-circle principle," that is, to assume the existence of impredicative totalities. But the very point of the standard interpretation is to assert the existence of every possible class independently of its characterization in language. Hence the standard interpretation would in fact provide a temptation of the very kind Dummett denied Frege.[30]

If Dummett credited Frege with the notion of arbitrary function, Nino Cocchiarella[31] credits him with no more and no less than the formulation of standard second-order logic:

> . . both Frege and Russell maintain that the logistic framework within which classical mathematics is to be represented must consist at least of a second order predicate logic where quantification is not only with respect to the role of singular terms, but to that of predicates as well. Indeed, it was Frege himself who first formulated and developed standard second order predicate logic, and he did so precisely as a framework within which classical mathematics was to be reduced to logic (*ibid.*, p. 198).

No reference to Frege's texts is provided by Cocchiarella.

---

[30] In fact, the only axioms of set theory which give us sufficient liberty in forming more comprehensive sets are the axiom of subsets and the power-set axiom, which, as is well-known, involve impredicativities.

[31] "Frege, Russell and Logicism: A Logical Reconstruction," in *Frege Synthesized*, L. Haaparanta and J. Hintikka, eds. (Boston Reidel, 1986), pp. 197–252

It is true that Frege used higher-order logic as a framework for reducing mathematics to logic, and that this can be done only by giving higher-order logic the standard interpretation. But there is no evidence in Cocchiarella or anywhere else that Frege actually perceived the need of the standard interpretation for this purpose or actually assumed it.

X. FREGE'S CRITICISMS OF "ABSTRACTION"

What made possible the conception of an arbitrary set was the gradual disentanglement of the notion of set from intensional ingredients such as concepts, properties, etc., and the definition of sethood in an alternative way. One such alternative way attempted by some foundationalists consisted in thinking of classes as mere collections of objects obtained by "abstraction." For instance, abstraction was typically the procedure through which Cantor claimed he could reach his notion of cardinal number which he defined as "the general concept which arises from the set M when we abstract both from the nature of the various elements m and from the order in which they are given" (*op. cit.*, p. 282). But abstraction was not, as such, limited only to the explanation of what numbers are but could also be used to explain what classes are. Thus, Edmund Husserl[32] argued that a collection (totality) consists of contents that are to be thought without any relation or connection between them.

Frege criticized consistently and constantly other logicians' and mathematicians' attempts to introduce arbitrary classes by abstraction. For instance, he criticized Husserl for mistakenly defending a notion of totality in which "the contents are merely supposed to be thought or presented together, without any relation or connection whatever being presented between them."[33] For his part, Frege confesses that he is unable "to disregard the connection without losing the whole" (*ibid.*).

Frege's forceful criticisms of abstraction have usually been taken to be merely a part of his overall campaign against psychologism in logic and mathematics.[34] This is a dangerous oversimplification. Frege's arguments, no matter how they were formulated, had a direct impact on the understanding of his substantial views of logic.

---

[32] *Philosophie der Arithmetik: psychologische und logische Untersuchung, Vol. I* (Halle a.d. Saale: Pfeffer, 1891), p. 79

[33] "Review of Dr. Husserl's *Philosophie der Arithmetik,*" *Mind,* LXXX (1972), p 329.

[34] For some of Frege's criticisms against psychologism, see *ibid.*, pp. 321–37; *Foundations,* p 63; *Collected Papers in Mathematics, Logic and Philosophy,* Brian McGuinness et al., eds. (Cambridge· Blackwell, 1984), pp. 178–81.

They are, in effect, arguments for the primacy of concepts over their potential extensions. Hence they have an indirect impact also on his views on higher-order logic. They show that Frege did not appreciate the notion of arbitrary set, and they show once more how unwilling he was to accept what later was to be called the standard interpretation.

XI THE TWO FACETS OF EXTENSIONALITY AND THE PROBLEMS OF SET THEORY

Mathematically and logically, the most important target of Frege's criticisms of abstraction was undoubtedly Cantor. A delicate touch is needed, however, in discussing the conceptual history of set theory and its relations to Frege's ideas. Earlier in this paper, we pointed out the deep kinship between the standard ("extensional") interpretation of higher-order logics and the intuitions on which set theory was originally based. This standardness, we pointed out, was lost in the course of the development of the contemporary (first-order) axiomatic set theory.

Yet a perspective that at first sight might appear to be diametrically opposite to ours has been presented and ably argued for by Hallett (*op. cit.*). In his story, set theory was originally afflicted with a serious case of intensionalism. This malady was gradually cured in the course of the development of axiomatic set theory, which slowly led to the all-too-familiar cumulative conception of sets. According to Hallett, Frege was on the side of extensionalist angels, and not the bogeyman that he may look like in our scenario. Indeed, Hallett (*op. cit.*, p. 46) brackets Frege in so many words with those nineteenth-century logicians and mathematicians, exemplified by Boole and Dedekind, who "operated with arbitrary extensions."

Furthermore, according to Hallett, the liberation of set theory from the dominance of the intensionalist conception was a painful and slow process. This can be seen from the fact that ingredients of the intensionalist conception are plenty in Cantor and still, surprisingly, in Zermelo. For instance, in early Cantor a set was defined as a "totality of elements which can be united to a whole through a law."[35] Also, in Zermelo's formulation of the subset (separation) axiom, what yields the subset is a "property."[36] It is only with Skolem that set theory was definitively purged of intensional entities like properties, which were replaced by predicates of the given language.[37]

---

[35] "Über unendliche, lineare Punktmannigfaltigkeiten 5," in *op. cit*

[36] See Zermelo, p. 183; and Hallett, p 267.

[37] A more detailed description of Skolem's ideas is contained in Hallett, p 281

It is nevertheless easy to reconcile Hallett's perspective with ours. What has happened is that Hallett tells only a half of the full story, whereas we are in this paper focusing on the other half. We have been asking: Does there exist, according to Frege or whoever else, a concept for each potential extension? Hallett is restricting his attention to an entirely different and in fact reverse question: Does there exist an extension for each concept? As was pointed out above, both questions have sometimes been discussed under the undifferentiated heading of "extensionalism," and even assimilated to each other.

In the course of the actual development of set theory, it was the second problem that was gradually forced upon mathematicians and logicians. This happened largely as a response to the paradoxes of set theory, which showed the inconsistency of certain kinds of assumptions of the existence of sets (corresponding to apparently irreproachable concepts). The result was a preoccupation with the conditions of the existence of extensions, given the corresponding concepts. What Hallett argues is that the development of the cumulative concept of set was looked upon as a process of liberalization of such existence conditions. We have no quarrel with his argument.

Yet, as far as the other problem is concerned, the upshot of the development of axiomatic set theory was to give up the standard interpretation and hence to give up Cantor's original vision of set theory. Even set theorists like von Neumann, who were beginning to see the distance between Zermelo's axiomatic system and its Cantorian ancestor and who tried to construct a richer set-theoretical universe, could never return to Cantor's conception of arbitrary set. This fact is pointed out by A. Fraenkel and Y. Bar-Hillel[58] when they discuss von Neumann's system:

> . . . The characteristic property of a function, namely being determined by the course of its values, is expressed by an appropriate axiom, . . . it states that functions with the same course of values are equal—but not that to each course of values a function is assigned, ***which would mean a return to Cantor's conception of an arbitrary set*** (*ibid.*, p. 99; italics added).

Thus, it is important to realize that Hallett's is not the only relevant perspective here nor, we believe, the most important one. What happens in type-theoretical higher-order logic is that the Gordian knot of paradoxes is cut open once and for all, not unraveled by means of sophisticated axioms. The type distinctions guarantee that the existence of the appropriate extension for each concept of the same type can never cause any problems. The entire attention switches to the problem of capturing those potential extensions which prima facie are not extensions of any one concept. Probably the simplest way of accomplishing this task is to adopt a nonstandard interpretation. Piecemeal adoption of new first-order (or nonstandardly interpreted higher-order) axioms will not make essential difference. What this means for someone like Frege (for whom sets exists only as extensions of the corresponding concepts) is that an important problem here is to guarantee the existence of suitable concepts.

---

[58] *Foundations of Set Theory* (Amsterdam: North Holland, 1958), esp. ch. II, sects 6–7.

XII FREGE AND THE TWO FACETS OF EXTENSIONALITY

In the light of these observations, you can now see that Hallett's perspective is perfectly compatible with ours, as long as it is understood that he is dealing with only one kind of problem in the foundations of set theory. Indeed, we can now appreciate better the discussions in which Frege was involved.

Hallett is entirely right in emphasizing that Frege wanted to assume an extension (course of values) for each and every function (in Frege's sense), including every concept. In fact, that assumption is built into the very notation of the *Grundgesetze*. This assumption was dragged out to the open and criticized by Cantor, who declares it "an unhappy idea":

> ... to take "the extension of a concept" as the foundation of the number concept. He [Frege] overlooks the fact that in general the "extension of a concept" is something quantitatively undetermined. Only in certain cases is the "extension of a concept" quantitatively determined; then it surely has, if it is finite, a definite natural number, and if infinite, a definite power [Mächtigkeit, i.e., cardinality]. For such quantitative determination of the "extension of a concept", the concepts "number" and "power" must previously be already given from somewhere else, and it is a reversal of the proper order when one undertakes to base the latter concepts on the concept "extension of a concept" (*op. cit.*, p. 440).

Here we see both the fact that Frege was indeed on the side of the "extensionalists" in the issue on which Hallett is concentrating. At the same time, the quote supports our thesis of the priority of concepts for Frege, in contradistinction to the priority of classes (extensions) for Cantor.

But if so, is there any trace in Frege of the real problems he has to face? Does he realize that, if he prioritizes concepts over their extensions, he has to provide enough concepts (and thereby enough sets) to satisfy the existential needs of actual mathematical theorizing, among other things to satisfy the needs of set theory? At the first sight, Frege does not evince any awareness of such problems. In a deeper sense, however, there are signs of a such awareness in Frege. Indeed, the very priority of concepts over their extensions provides a clue. For one does not normally speak of assuming the existence of a concept but of *defining* the concept. Frege clearly thinks that a definition can bring new sets into a logician's reasoning. This is seen from his discussion in *Grundlagen,* section 88, of the analytic versus synthetic distinction and its relation to definitions. The upshot of this discussion is that what Frege calls definitions can extend our knowledge and hence be synthetic in Kant's sense. Admittedly, Frege says that the consequences of such definitions can be proved by purely logical means and hence are analytic in his sense. But here we must remember that for Frege logical truths are simply the most general truths about this world. Therefore, they presumably can involve existential truths. It seems to us that definitions are for Frege a way of eliciting the existential implications of what for him were purely logical truths.

This is strongly suggested in Frege's own illustration:

> If we represent the concepts (or their extensions) by figures or areas in a plane, then the concept defined by a simple list of characteristics corresponds to the area common to all the areas representing the defining characteristics; it is enclosed by segments of their boundary lines. With a definition like this, therefore, what we do—in terms of our illustration—is to use the lines already given in a new way for the purpose of demarcating an area . . . But the most fruitful type of definition is a matter of drawing boundary lines that were not previously given at all. What we shall be able to infer from it, can be inspected in advance; here, we are not simply taking out of the box again what we have just put into it (*Foundations,* pp. 100–1).

But if so, we can predict what the focus must have been for Frege's doubts about the way in which Cantor and others were introducing the sets they needed in their theories. The prediction is that Frege must have been worried about the way these logicians and gentlemen defined their basic concepts. This would-be prediction is indeed fulfilled. For Frege's criticism of Cantor is leveled prominently at his use of definitions. This criticism is found in an especially sharp formulation in Frege's originally unpublished review of Cantor.[39] It involves prominently an attack on Cantor's notion of abstraction, thus indicating its tie to Frege's rejection of the standard interpretation.

The relation of Frege's notion of definition to existence assumptions deserves a fuller examination than we can give it here. But even without such a further investigation, his discussion of other logicians' and set theorists' definitions is suggestive of this half-awareness of the crucial (for him) problem of the existence of concepts. Unfortunately, this problem was swept out of the focus of Frege's attention by the more immediately urgent problem of the existence of extensions for (so to speak already existing) concepts.

XIII. "CONSISTENCY DOES NOT IMPLY EXISTENCE" AND THE STANDARD VERSUS NONSTANDARD DISTINCTION

There are other ways of seeing that Frege was inextricably committed to a nonstandard interpretation. On several occasions, Frege bases his arguments on the principle that we must not infer existence from consistency (freedom of contradiction). For instance, he writes:

> Suppose we have proved that in an equilateral, rectangular triangle the square on the hypotenuse is twice as large as the square on one of the sides, which is easier than proving Pythagoras' theorem in general. We can now conclude that the proposition does not contain a contradiction either within itself or in relation to the geometrical axioms. But can we conclude further: therefore, Pythagoras' theorem is true? I cannot accept such a method of inference from lack of contradiction to truth.[40]

The principle that Frege rules out appears to be what would be needed to guarantee that there is a concept for any class (possible extension). What this possibility means in this case is that, if a concept should somehow be introduced which actually has this class as its extension, no contradiction would come about. Hence the forbidden principle could guarantee the existence of the "missing" concept.

Admittedly, Frege does not in so many words apply his principle "existence does not follow from consistency" to deny that there is a concept for each class, having this class as this extension. But the variety of different cases to which he does apply his principle strongly suggests that he was committed to applying it also so as to rule out the standard interpretation.

---

[39] "Entwurf zu einer Besprechung von Cantors Gesammelten Abhandlungen zur Lehre vom Transfiniten," in *Gottlob Frege: Nachgelassene Schriften*, H. Hermes, F. Kambartel, and F. Kaulbach, eds. (Hamburg: Meiner, 1969), pp. 76–80.

[40] *Correspondence*, p. 48.

In the *Grundlagen,* Frege admittedly uses his principle only to criticize H. Hankel's somewhat cavalier procedure in extending the concept of number by using as his only firm guideline the preservation of the main laws governing numbers. It might seem that no sweeping conclusions can be drawn from Frege's facile triumph. But surely Frege could and should have tried to criticize more significant thinkers than Hankel in the same way. What Hankel does is not so entirely different from Dedekind's procedure. Dedekind, too, is in effect relying on the consistency of the result when he introduces a real to correspond to any cut, even when it does not correspond to a real number. And if Frege had so extended his criticisms to Dedekind, their force would once again come very close to the denial of the standard interpretation.

XIV. FREGE DISAVOWS THE CONCEPT OF ARBITRARY FUNCTION

The most convincing evidence for ascribing a nonstandard interpretation to Frege is mediated by the connection pointed out earlier between the standard interpretation and the idea of arbitrary function. Frege's conception of function is shown not only by what he says about it but also how he applies the concept of function. In his early paper, "Methods of Calculation Based on an Extension of the Concept of Quality,"[41] Frege claims to have proved "some general propositions . . . valid for all functions and functional systems" (*ibid.,* p. 92). Yet in actually proving his results, Frege is freely assuming that the functions he is dealing with are differentiable (*ibid.,* p. 60). In other words, Frege's "all functions" are not all arbitrary functions in the usual sense of the word. Frege's conception of function is obviously light years away from the concept of arbitrary function in the sense that is related to the standard interpretation.

Even more striking than the narrowness of Frege's working concept of a function is his obliviousness to the problems afflicting this notion. As was noted earlier, discussions concerning the general concept of function were perhaps the most important ingredient in the lively foundational discussions among mathematicians in Frege's lifetime. Hence when Frege takes up in so many words the critical question "What is a function?"[42] one might therefore expect some awareness of the actual cases related to mathematical research in such areas as Fourier analysis. This expectation is not fulfilled. Instead, we find in effect an unmistakeable rejection of the concept of arbitrary function in the relevant sense of the word. Frege formulates, modifying E. Czuber's definition, the idea of an arbitrary function as follows:

---

[41] In *Collected Papers,* pp. 50–92

[42] *Ibid.,* pp. 285–92

> With every number of an x-range there is correlated a number. I call the totality of these numbers the y-range (*ibid.*, p. 289).

Frege calls our attention, quite correctly, to the notion of correlation involved in this concept of function:

> The heart of the matter really lies in quite a different place, viz. hidden in the word 'correlated' (*ibid.*).

He notes that his opponent leaves the nature of this correlation completely open, quoting Czuber as saying that

> The above definition involves no assertion as to the law of correlation, which is indicated in the most general way by the characteristic f; this can be set up in the most various ways (*ibid.*).

Frege immediately focuses on the idea of "law":

> Correlation, then, takes place according to a law, and different laws of this sort can be thought of. In that case, the expression *y* is a function of *x* has no sense, unless it is completed by the law of correlation (*ibid.*).

But how can we do that? Frege answers:

> Our general way of expressing such a law of correlation is an equation in which the letter 'y' stands on the left side whereas on the right there appears a mathematical expression consisting of numerals, and the letter 'x', e.g. '$y = x^2 + 3x$' (*ibid.*).

Here Frege is not speaking of "correlations" as being exhibited by objectively existing entities out there in the Platonic realm of numbers, called functions, but by actual mathematical equations expressed in a suitable notation.

This by itself does not yet imply that Frege rejected the idea of arbitrary function. For it leaves open the possibility that laws definable by equations do not exhaust the range of all functions he is countenancing. Helpfully, Frege nevertheless clarifies his position immediately. He continues:

> Functions have indeed been defined as being such mathematical expressions. In recent times this concept has been found too narrow.

> However, this difficulty could easily be avoided by introducing new signs into the symbolic language of arithmetic (*ibid.*, p. 290).

This is a mind-boggling statement. It shows Frege's mind at work. It shows that there is no niche in his world for the notion of an arbitrary function in the sense in which Euler, Czuber, and others were thinking of this notion. For to think that the range of such arbitrary functions could be covered by introducing new signs into the notation of arithmetic is not to understand this notion. Hence we can safely conclude that, in the logic to which Frege tried to reduce all mathematics (except geometry), there is no room for the idea of an arbitrary "function-in-extension" and hence no room for the standard interpretation.

It is not that Frege accepts uncritically the characterization of the largest relevant class of functions by means of equations. Indeed, he immediately goes on to criticize the quoted characterization. But this criticism has nothing to do with the narrowness of the class of functions so definable. It deals with the fact that the proposed characterization obscures the "distinction between sign and the thing signified," and hence obscures the objective existence of functions independently of any particular way they are formally expressed.

What is interesting about this justified criticism by Frege is that its upshot obviously has nothing to do with the range of functions that exist or do not exist. Problems of that kind, Frege has already proclaimed, "can easily be avoided by introducing new signs into the symbolic language of arithmetic." Hence Dummett and others are simply wrong if they think that the objective existence of functions, or even their mode of objective existence as correlations of objects with objects, prejudices Frege in favor of the standard interpretation.

At the same time, the latest displayed quote from Frege boggles the mind in that it demonstrates Frege's alienation from the actual working problems in the foundations of mathematics of his time. Frege is in effect telling Dirichlet, Weierstrass, Dedekind, H. Poincáre, et alia that the problems they have been agonizing about could easily by solved by introducing new signs into the notation of arithmetic. It is perhaps not surprising that Frege's foundational ideas did not attract more attention among the practicing foundationalists than they in fact did.

XV. FREGE'S REDUCTIONIST PROGRAM IS A NONSTARTER

What have we accomplished so far? And why have we done what we have done? Why worry about whether Frege accepted the standard or nonstandard interpretation of higher-order logic? After all, the distinction was articulated much later, and hence Frege could scarcely have expected to pay much attention to this distinction, which might seem rather pedantic. Yet there is a massive reason for raising the question we have raised. *The evaluation of Frege's entire life work depends crucially on the answer.* Frege tried to reduce mathematics to logic. In the process, he used crucially higher-order quantifiers. For instance, he defined "The number which belongs to the concept F is ∞_1" as "there exists a relation which correlates one-to-one the objects falling under the concept F with the finite Numbers" (*Foundations,* p. 98). Despite the fact that Frege claimed that his terminology "diverges to some extent" from Cantor's, in the *Grundlagen* he welcomes Cantor's achievements because "they have led to the construction of a purely arithmetical route to higher transfinite Numbers (powers)" (*ibid.,* p. 96). It has been emphasized in this paper, not for the first time, that the divergence between Frege and Cantor is much deeper than merely a terminological one. The divergence spreads from two radically different conceptions of a set, one that ties sets to concepts and properties, the other that generates sets freely through operations acting on the members of some already existing sets. Likewise, it makes more than a terminological difference whether one defines a one-to-one correspondence between two sets as any arbitrary one-to-one function in extension mapping them on each other, as Cantor envisaged it, or as a similar function picked up by a concept, as Frege in effect did.

Also, when Dedekind characterized the real numbers in effect by means of the cut principle, which says that every bounded (from the above) set of reals has a least upper bound, this characterization is a categorical one only if the sets involved are arbitrary ones and not only such sets as are courses of values of concepts expressible in the language of arithmetic. Thus, there is a deep sense in which Frege's system is not adequate for interpreting results in contemporary set theory and mathematical theorizing, for instance in real analysis.

XVI. FREGE'S PLACE IN HISTORY

Even more generally, serious questions must be raised in the light of our results and of other results about Frege's stature as a philosopher of mathematics and as an analyst of the actual working foundations of mathematics. Frege created our received conception of logic, not just in the sense of conceiving the general idea but in the sense of actually implementing it. This is a lasting achievement. Yet it turns out that his creation, the customary idea of first-order logic, is in certain respects too restrictive and perhaps even arbitrarily restricted. As Hintikka[43] has recently shown, Frege and others have arbitrarily excluded quantifier independence from their logic. As we have seen, Frege's failure to grasp the idea of the standard interpretation of higher-order logic turns his entire foundational project into a hopeless daydream. Elsewhere, Hintikka[44] has pointed out how totally blind Frege was to what must be considered the most significant line of development in modern foundational studies, the gradual systematization of the model-theoretic viewpoint. In particular, Frege totally failed to understand the early uses of this viewpoint by David Hilbert, especially in Hilbert's work in the axiomatic foundations of geometry.

As Hilbert's work was easily the most influential kind of foundational work among mathematicians and scientists in the early decades of this century, this means that Frege was totally blind to one of the most important lines of thought in the foundations of mathematics in this century. What all of this adds up to is that Frege failed to relate his own work to much, not to say most, of the foundational work that really mattered to actual mathematical research and to mathematicians' own efforts to clarify the fundamentals of their own discipline. It seems to us unfortunate that philosophers habitually go to Frege for their problems and for the concepts that could help us to cope with them. Frege was far too myopic to be a fruitful source of concepts, ideas, and problems. It is likely to turn out that some of his contemporaries can offer far better ideas for future work in the foundations of mathematics than Frege did.

JAAKKO HINTIKKA

Boston University

GABRIEL SANDU

University of Helsinki

---

[43] "What is Elementary Logic?" *Proceedings of the Boston Colloquium in Philosophy and History of Science* (Dordrecht. Kluwer, forthcoming).

[44] "On the Development of the Model-Theoretical Tradition in Logical Theory," *Synthese*, LXXVII (1988): 1–36.

BOOK REVIEWS

Hegel's Ethical Thought. ALLEN W. WOOD. New York: Cambridge University Press, 1990. xxi, 293 p. Cloth $54.95, paper $17.95.*

Allen Wood's latest book is the best existing treatment of Hegel's ethical thought by a philosopher in the Anglo-American tradition. It is a sympathetic and comprehensive reconstruction of the philosophical foundations that underly Hegel's views on society, politics, and morality. The philosophical foundations investigated here are not the speculative metaphysics of Hegel's *Logic* (which Wood regards as a dead enterprise) but Hegel's ethical theory: his understanding of the good and its relation to reason, freedom, and human happiness. Wood's clarity and philosophical expertise make this book a highly valuable resource for students of Hegel, both beginning and advanced, as well as for analytic philosophers who want a comprehensible account of Hegel's ethical thought which is free of obfuscating jargon. Wood successfully demonstrates both that Hegel has an ethical theory and that it is a powerful view worthy of our attention.

Wood classifies Hegel's normative ethical theory as a "self-actualization theory," which he distinguishes from both deontological and teleological positions. An ethical theory based on self-actualization regards as good those actions and ends which are espoused by rational subjects attempting to realize their conceptions of their own essential nature. In Hegel's theory the subject's essence that is to be actualized is freedom or, more accurately, freedom in its various configurations. The latter modification is necessary, since freedom for Hegel assumes a variety of forms, and since the ordering of these forms plays a crucial role in determining the architectonic of his theory. The principles of the three main parts of Hegel's theory—abstract right, morality, and ethical life (*Sittlichkeit*)—are derived from three distinct (but complementary) ways in which subjects of the modern world understand themselves as free. Because Hegel's position is grounded in a single fundamental good (freedom), it comes close to being a teleological theory that posits freedom as the telos of ethical action. Yet Wood rejects the teleological label because it mistakenly suggests that ethical action can be understood as a means to some end external to the action. On Wood's reading of Hegel, actions have ethical value not because they are instrumental

---

* I would like to thank John Rawls for his comments on an earlier version of this review

0022-362X/92/8906/316–20

to the achievement of some independent good, but because acting through the forms of freedom is the way in which a subject objectifies, or gives reality to, its conception of itself as free. (Owning property, for example, has ethical significance not because it has an independently valued end as its consequence, but because the will's relationship to its property satisfies a conception the modern subject has of itself as a free chooser of ends.) Who, then, is the subject of self-actualization: Spirit or human individuals? To this question Wood replies that it is *both,* and that Hegel's emphasis on the notion of Spirit does not diminish the importance of individuals in his theory, since the self-actualization of Spirit occurs only through the self-actualization of individuals. We should wonder, however, whether this fact alone warrants Wood's conclusion that "Hegel's ethical theory is after all *founded* on a conception of individual human beings and their self-actualization" (21, emphasis added). I return to this issue below in my discussion of Hegel's relation to liberalism. Although Wood's book contains many ideas worthy of careful attention (including an excellent treatment of freedom, happiness, and their relationship), I shall restrict myself to commenting on two problems raised by his interpretation.

THE EMPTINESS OF THE CATEGORICAL IMPERATIVE

The best known of Hegel's ethical doctrines is his charge that the basic principle of Kantian morality, the categorical imperative (CI), is empty and therefore unable to serve as a criterion for what is morally right and wrong. This view has received a good deal of attention from commentators, most likely because there is something intuitively compelling in the claim that the formal nature of Kant's pure practical reason makes it unable to supply the notion of moral duty with determinate content. Yet even sympathetic interpreters have had a difficult time making sense of this critique, for the specific arguments Hegel marshalls in support of it (in his famous examples of aiding the poor and returning the deposit of the deceased) seem to rest upon a gross misunderstanding of Kant's procedure for testing an agent's maxims. Wood makes a provocative suggestion as to how Hegel's critique must be understood if we are to be able to take it seriously. His interpretive strategy is to distinguish between two versions of the emptiness charge, the first of which is said to fail as a criticism of Kant, whereas the second allegedly succeeds in pointing out an important defect in Kant's moral philosophy.

The first version of the emptiness charge, the one most commonly attributed to Hegel, claims that the CI (especially in its first formulation in the *Groundwork*) cannot distinguish between permissible and impermissible maxims: from one perspective all maxims pass the test of universalizability; from another all maxims fail. Wood claims not only that Hegel's particular examples fail to support his charge but that the critique itself is untenable. By reconstructing Kant's example of the unreturned deposit, Wood argues that the CI procedure, properly understood, is able to rule out some maxims as impermissible. Although the procedure does run into serious problems (157, 160), emptiness in the sense alleged by Hegel is not one of them. The second version of the emptiness charge, which Wood defends as a cogent critique, is directed not at Kant's formulations of the CI but at something more fundamental: the philosophical position from which Kant attempts to derive and defend this basic moral principle. On this reading, Hegel's claim is that no contentful ethical principles can be derived from the abstract notions of reason and subjectivity which ground Kant's moral philosophy. Wood fleshes out this claim in an attack on Kant's attempt to arrive at a formulation of the supreme principle of morality starting merely from the concept of an objective practical law, or a categorical imperative. Wood's critique of Kant's derivation of the CI is plausible and makes good philosophical sense of Hegel's charge of emptiness. Moreover, his interpretation suggests a way of understanding Hegel's well-known examples which does not attribute to Hegel a simplistic distortion of Kant's position. Their point is not to demonstrate the emptiness of the CI as it is in fact formulated by Kant but rather to show the emptiness of the only criterion (logical noncontradictoriness) available to Kant, given the conception of reason from which he intends to derive his basic moral principle.

This interpretation raises an important question: How does Hegel's own position enable him to avoid the problem of emptiness that plagues Kant? Hegel's solution presumably lies in his distinctive notion of ethical life and its conception of subjectivity, which defines agents in terms of the roles they occupy in the basic institutions of a rational society. Although Wood offers a number of helpful suggestions as to how the perspective of ethical life is supposed to supply morality with content, two important unclarities remain in his account of Hegel's mature conception of morality and its relationship to a theory of social institutions: First, what precisely are the resources that Hegel's own moral standpoint has at its disposal? Does Hegelian morality simply do away with the kind of practical reasoning involved in Kant's CI procedure, or does it retain something like the universalizability criterion which, when coupled with a commitment to the rational social institutions of ethical life, becomes capable of discriminating between permissible and impermissible maxims? Second, how are we to envision the division of labor by means of which the principles of morality collaborate with those of ethical life to yield a determinate set of duties? Is it the case, as Wood sometimes suggests (173, 210–1), that the standpoint of morality gives rise itself to a set of *moral* duties (e.g., acts of charity to strangers), which are supplemented by ethical life in the sense that the latter provides an additional set of distinctively *ethical* duties (e.g., family obligations, respect for civil law)? On this picture, morality is incomplete without the institutions of ethical life but not, strictly speaking, empty. Or is it rather (136, 172) that the moral standpoint is indeed empty in the sense that the categories it employs (the "rights" and "well-being" of others) lack all content when considered in abstraction from ethical institutions? This reading (which seems to accord best with §§135R and 148R of the *Philosophy of Right* (PhR)) holds that morality on its own "provides us with no determinate criteria of action" and that the latter can be "supplied only by the more concrete social relationships of ethical life" (136). And, if the latter is indeed Hegel's view, can one defend his thesis concerning the emptiness of the purely moral standpoint, given Wood's argument in response to the first version of the emptiness charge that Kant's CI can provide contentful criteria for right action?

HEGEL AND LIBERALISM

One topic that runs throughout this book is Hegel's relationship to liberalism. Wood provides a valuable corrective to current interpretations by emphasizing those aspects of Hegel's thought which link him to the liberal and Enlightenment traditions and distance him from contemporary communitarianism. On Wood's interpretation, Hegel espouses political institutions that are "pretty much" those defended by liberals but rejects the philosophical grounds on which liberals justify those institutions. Although Wood's treatment of Hegel's relationship to liberal theory is generally illuminating, I shall focus on one aspect of his discussion that seems to me problematic: the question of whether, and in what sense, a collective good (here the good of the state) is reducible to the good of the individuals who compose it. According to Wood, Hegel subscribes (along with liberals) to "the virtual tautology that collective goods have value because they have value *for individuals*" (259). Hegel is said to diverge from liberal theories in asserting that "collective goods have a value in themselves not reducible to the *private* good of individuals" (258, emphasis added). Presumably, Wood is pointing here to Hegel's idea that the value of the state resides in something beyond its function of enabling citizens to attain ends they possess as private individuals in abstraction from their political relations to others.

Participation in the political sphere is more than merely a means for achieving private goods; as the activity through which citizens determine their own social existence, it is valuable in itself. But this aspect of Hegel's position does not in fact distinguish him from much of the liberal tradition. To cite just two examples, neither Rousseau nor John Rawls would deny that the value of the political realm extends beyond its instrumental value for the attainment of private ends. If there is a difference here, I believe it is more accurately located in Hegel's insistence that individuals achieve their ***highest*** good (as practical beings) in political activity.

An even more fundamental difference comes to light if we reconsider Wood's claim that for Hegel collective goods have value because they have value for individuals. It is certainly Hegel's view that a necessary condition of a good's having collective value is that it be valuable for individuals. But it does not follow from this that a collective good has value only because of its value for individuals. This point is significant, because it casts doubt on Wood's conclusion that Hegel's ethical theory is ultimately *founded* on the notion of individual self-actualization. Indeed, this seems to me a misleading characterization of Hegel's view, for his conception of individual self-actualization is derivative of his understanding of the self-actualization of Spirit. Individual self-actualization is an important notion for Hegel, but it is understood in such a way that individuals realize their essence not *as individuals* but as members of a collective, i.e., as vehicles of Spirit. Individuals achieve their highest freedom by participating in and identifying with a social whole that is itself a complete and self-sufficient good. Furthermore, Hegel's *organic* conception of the social whole and the collective good requires that there be a substantial differentiation, even an ethical hierarchy, among members of the state which violates the basic liberal view of citizens as morally equal individuals (PhR, §308R). The peculiar nature of Hegel's understanding of individual self-actualization is brought out best in his treatment of social groups that are denied full participation as citizens. The peasant, for example, occupies a role in civil society that is incompatible with the full development of the human capacities for rationality and individual self-determination. Yet the individual peasant realizes his essence as free in the rational state. The crucial point is that he does so, not because he achieves autonomy as an individual, but because he is an integral part of a community that, as a whole, is self-sufficient and therefore undetermined by anything other than itself.

FREDERICK NEUHOUSER

Harvard University

Realism in Mathematics. PENELOPE MADDY. New York: Oxford University Press, 1990. ix, 204 p. Cloth $39.95.*

Maddy defends a form of realism intended to be true to mathematical practice against this objection: on a realist account, mathematics would be about abstract objects. Since we cannot interact with abstract objects, mathematical knowledge would then be impossible. She seeks to explain how mathematicians ordinarily obtain mathematical knowledge in a way that is consonant with and makes use of the best contemporary scientific theory of whatever turns out to be relevant.

Some readers may be puzzled at this point. What Maddy calls the Quine/Putnam indispensability argument explains the possibility of mathematical knowledge: mathematical objects are indispensable to our best theory of the natural world, and that theory yields mathematical knowledge. But mathematicians do not normally make use of extramathematical scientific theories to obtain mathematical knowledge. The Quine/Putnam indispensability argument does not explain how mathematical knowledge is actually obtained.

The obvious method of obtaining mathematical knowledge—proving theorems—is no help: a proof shows that if the axioms are true then so is a theorem. But how do we know that the axioms are true?[1] Mathematicians seem to use a faculty of mathematical intuition that plays a role for mathematics analogous to that sense perception plays for physics. Maddy's problem is (at least in part) that of giving a naturalistic explanation both of how mathematical intuition is possible and of the role it plays in mathematical knowledge.

We have been discussing mathematics generally, but, "to bring the project down to more manageable size," Maddy only presents her solution for one branch of mathematics—set theory. Set theory poses a particularly interesting challenge for any theory of how we come to accept axioms, since its practitioners actively seek out new axioms.

Maddy's solution to the problem of mathematical intuition is simple: we see some small simple sets in the same kind of way that we see some medium-sized physical objects. The mathematician's intuitions about sets are essentially identical in kind to the natural scientist's intuitions about physical objects: both are products of the ap-

---

* I would like to thank Bernard Berofsky and Sidney Morgenbesser for useful discussions.

[1] There is a long tradition of denying the importance or even the coherence of this question, and a long tradition of arguing that that denial is untenable. Maddy reviews the situation and concludes that the denial is untenable.

0022-362X/92/8906/321–6

paratus we use to perceive the sets and physical objects. The relevant sense of "seeing" is that in which seeing a medium-sized physical object enables the person who sees it to gain appropriate perceptual beliefs about it. The work of Piaget and others has taught us that the ability to see a medium-sized physical object is the result of a long and complex developmental process. According to Hebb's neurophysiological theory, this developmental process involves successive alterations in neural structuring in the brain which culminate in the acquisition of a complex cell assembly that serves as a physical-object detector. To be capable of seeing a physical object is just to have such a physical-object detector. Our neuropsychological theories will surely change, but there is reason to believe that whatever form our account of vision comes to take will allow Maddy's arguments to go through without essential modification. Maddy tries to show that one can give an account of how we perceive sets similar to the one for physical objects.

Take Maddy's example of Steve's opening a carton of eggs and seeing that the set of eggs in it has three members. By analogy to perception of physical objects, she must show that there is such a set, that Steve acquired perceptual beliefs about it, and that the set participated in the acquisition of those beliefs in the right way. The belief that there are three eggs is perceptual, not inferential, as is indicated, for example, by reaction-time studies. But why should we take the belief to be about a set? As Frege taught us, the stuff that makes up the three eggs also makes up some huge number of molecules. Whatever three is a property of, it is not a property of the underlying matter. The usual candidates include aggregates, concepts, classes, and sets. Quine/Putnamism enters in: sets play a role in our best theory of the world, while the other entities named do not. Thus, according to Maddy, Steve has acquired a perceptual belief about a set. But has he *seen* a set?

The remaining point is that the set of eggs has participated in the acquisition of Steve's beliefs in the right way. Maddy adopts Grice's method of finessing the problem of what "the right way" is: "the thing perceived must play the same sort of role in the causation of the perceiver's perceptual state as my hand plays in the generation of my belief that there is a hand before me when I look at it in good light." The hand plays the following role: light that has interacted with an appropriate part of it is focused on the retina, and that causes appropriate cell assemblies to be stimulated in an appropriate sequence. To see a set, Steve must have appropriate cell assemblies that developed in the same sort of way as did the ones that play a role in seeing a hand in front of your face. Maddy tells a satisfactory story involving Piaget on sets and a Hebbian neural "set-detector." In addition, light that has interacted with an appropriate part of the set must have been focused on Steve's retina. But the eggs, which are what the light interacted with, are not parts of the set. Thus, despite Maddy's arguments, Steve did not see the set. Here is another example: according to Maddy, "the most natural way of understanding" the perceptual belief "here are three pairs of shoes" involves a set of sets—the set whose members are three sets, each of which contains two shoes as members. Even though the shoes are members of members of the set, they are not parts of the set. There is a more direct argument to the same conclusion: sets are abstract objects, and so they cannot have causal interactions.

Maddy established that Steve's belief that there are three eggs is a perceptual belief about a set. Thus, since the last criterion for perception of a set fails, what we conclude is that Steve acquired a perceptual belief about the set without having perceived it. That is not paradoxical: when you pass a streetlight at night and see the shadow of someone carrying a knife, you acquire a perceptual belief that there is a mugger behind you, but you have not seen the mugger. If we acquire perceptual beliefs about sets, that suffices for Maddy's account of mathematical intuition. Since Maddy maintains in addition that we see sets, she grants that we can causally interact with them, and hence that sets are not abstract. The amended account avoids the need to grant that. One cannot perceive the empty set, and so Maddy owes us an account of why perceptual beliefs that a set has zero members are to be analyzed differently from perceptual beliefs that a set has one or more members. On the amended account, the empty set can be handled in the same way that other small sets can.

Maddy only exploits her argument to account for mathematical intuition, but it is likely that parallel arguments could be used to maintain that we perceive a wide variety of things ranging from statements to melodies to causes. Her argument can be fruitfully recast in a contrapositive form: skepticism about the possibility of giving a naturalistic account of seeing sets leads to parallel skepticism about the possibility of giving a naturalistic account of seeing medium-sized physical objects.

In the process of acquiring the concepts of set and physical object, those concepts become structured in such a way that some beliefs are inevitable while others are impossible. Those intuitive constraints on belief arise with the concepts themselves. The combination of Piaget and Hebb shows how the process might be naturalistically realized. Since those intuitions are obtained in virtue of the right kinds of interactions, we have, Maddy claims, some reason to believe they will be reliable.

Has Maddy succeeded in her task of defending "the mathematician's pre-philosophical attitude"? We must consider set theory and its history. Maddy gives an eminently readable account of the development of set theory from its inception through the 1980s which is carefully focused on philosophically relevant aspects. This part of the book is a *tour de force* of exposition that makes much important information accessible. I shall only discuss the axiom of choice here.

Cantor used the well-ordering principle—every "well-defined" set can be well ordered. Zermelo gave a proof (1904)[2] that used the assumption

> that even for an infinite totality of sets there are always mappings that associate with every set one of its elements. . .

That is Zermelo's version of the axiom of choice. Modern versions of the axiom drop the requirement that the totality be infinite. Nonetheless, the whole point is the infinite case: it is easily proved that finite totalities have such "mappings"—choice functions. Zermelo proposed that the axiom be accepted on the grounds that it is "intuitively evident and necessary for science . . ." (1908a). Baire, Borel, and Lebesgue opposed the axiom, arguing that it was intuitively implausible. The clash of intuitions prima facie casts doubt on the reliability of mathematical intuition. As Maddy argues in detail, however, the apparent unreliability of intuition in this case can be explained away as the result of mixing up reliable intuitions about two distinct types of entities—mathematical and logical collections. Intuition reliably indicates that the axiom of choice is true of mathematical collections, and mathematical developments show this to be the more useful notion of collection on which to focus. This account of the axiom of choice seems to me to be correct, by and large, and Maddy has done a real service in presenting it in detail.[3]

Maddy's history of the axiom of choice shows there are two different notions of collection about which we have differing intuitions. On her account of intuition, there must then be two different concepts of collection acquired as the result of different experiences. That conflicts with her account of the acquisition of the set concept, which only made room for one.

It may be possible to give an account of how theoretical considerations split the intuitive set concept into two. That would not suffice for Maddy: she must explain how we have two notions of collection about which we have differing intuitions, not just differing beliefs, and for her intuitions cannot arise as the result of theoretical considerations. Given Maddy's set-theoretic realism and her account of the origins of intuition, she must presumably show that we can see sets (the mathematical notion), and that we can see classes (the logical notion), and that we can tell whether we are seeing a set or a class, and that our differing experiences of sets and classes are adequate to give reliable intuitions for and against the axiom of choice, respectively. That seems a hopeless task.

---

[2] References are to the bibliography of the book under review.

[3] Something like it was suggested earlier by R. Bunn, *From the Calculus to Set Theory 1630–1910*, I. Grattan-Guinness, ed. (London: Duckworth, 1980), ch. 6.

We now consider whether our intuitions concerning mathematical collections count in favor of the axiom of choice. History is unequivocal: we have intuitions concerning mathematical collections which legitimately count in favor of the axiom. But how can that be the case?

We begin with Maddy's answer to a different question: How can experience of a limited number of things lead to legitimate expectations concerning anything beyond that number, and about which things should it lead to expectations? According to Maddy, it is an objective fact that things are carved up into natural kinds. Our experience of a limited number of things of the same natural kind leads to evidence concerning other things of that kind. Thus, if liquids form a natural kind, our experience that liquids we come into contact with are wet counts in favor of the expectation that liquids are wet.

Maddy takes sets to form a natural kind. That is how she explains why we should take the primitive concept of "something that has a number property" to be a concept of sets, not something else. It also figures in the explanation of how we know that the intuitions that arise in virtue of that concept of set are intuitions concerning sets.

Our experience that liquids are wet does not legitimately count in favor of the expectation that all liquids are wet: the liquids we come into contact with are aqueous, and not all liquids are. Our experience tells us nothing about whether nonaqueous liquids are wet. More exactly, if a chemist knows that not all liquids are aqueous, but that all of the liquids in daily experience are aqueous, then that chemist cannot legitimately expect that nonaqueous liquids are wet on the basis of daily experience. That is even more true if the chemist has reason to believe that aqueousness is relevant to wetness.

We return to the axiom of choice. All of the sets in our daily experience have choice functions—and that is itself a part of daily experience (let us just grant that for the sake of argument). Something like that must be what Maddy has in mind when she refers to "the intuitive evidence for choice." Unfortunately, we are in a liquid-aqueous liquid situation. If sets form a natural kind, then so do finite sets. All of the sets in our experience are finite. Thus, we do not have intuitive evidence that infinite sets have choice functions. Moreover, finiteness is relevant to having a choice function.

Zermelo probably fully understood the situation in 1904, when he formulated the axiom of choice. In 1908 (1908a) he makes that understanding explicit. Zermelo can no more rely on the type of intuition described by Maddy than our chemist could rely on daily experience, and yet Zermelo says that the axiom of choice is "self-evident." And Zermelo is the beginning of a long tradition that is dominant today. Maddy cannot deny the self-evidence of the axiom of choice on the basis of the controversies that have surrounded it, since she has successfully explained those controversies away. She must therefore admit that her account of set-theoretic intuition is incomplete. Since she wants to give a naturalistic account of intuition, and "the human mind is finite," that account will have to explain how we can in a warranted manner extrapolate from our experience of finite things into the infinite.

The gap in Maddy's story should not blind us to her accomplishments. She has provided the beginnings of a classification of the various justificatory practices of mathematicians, in so doing giving a philosophical analysis of the relevant history. And she has focused on one particularly troublesome justificatory practice of mathematicians: appeals to intuition. Even if Maddy's account of mathematical intuition is mistaken in some of its details, and even though that account is incomplete, it remains a counterexample to the claim that a naturalistic account of mathematical intuition is impossible. Her work has made it clear that the problem of giving an adequate account of mathematical intuition is one of the central problems in the philosophy of mathematics.

SHAUGHAN LAVINE

Columbia University

The Realm of Rights. JUDITH JARVIS THOMSON. Cambridge: Harvard University Press, 1990. viii, 383 p. Cloth $35.00.

Judith Jarvis Thomson is famous, or infamous, for her imaginative appeal to test cases in philosophical argument. In *The Realm of Rights* she defends and uses brilliantly her method of connecting our, or her, presystematic moral judgments about how someone

0022-362X/92/8906/326–9

may or ought to act under specified circumstances with very general theses in ethical theory. Her book is a gold mine of important hypotheses about what it is to have a right and what rights we have. Any moral philosopher who extracts the true gold and discards the few iron pyrites will be much the richer.

Thomson reconstructs Hohfeld's ideas of the fundamental legal conceptions and uses them in her analysis of our moral rights. Following him, she takes claims to be fundamental to rights and a claim of *X* against *Y* to be equivalent to a duty of *Y* to *X*. Precisely what kind of a behavioral constraint, then, is a duty of *Y* to *X*? According to Thomson, what is central is (with qualifications) that it is morally permissible for *Y* to infringe *X*'s claim if and only if *Y* would thereby produce a sufficiently large and appropriately distributed increment of good or advantage (175). Additionally, other things being equal, *Y* ought to seek a release from *X* in advance and *Y* ought to compensate *X* later for such harms or losses as *Y* causes *X* by any infringing of the claim (200–1).

This seems correct as far as it goes, for one may not infringe a claim merely to produce a marginal increment of value when all those affected by one's act are counted. Still, more needs to be said. How can one explain the relational aspect of *X*'s claim *against Y* consisting in *Y*'s duty *to X*? The only hint I can find is when Thomson suggests that a legal duty might consist in the legal system's attaching a penalty to "*Y*'s doing such and such to *X*" (71). Now, possibly a relative duty is owed to the patient, the party who would be acted on in any violation of that duty. Hohfeld, however, suggests that a relative legal duty is owed to either the party benefited by its performance or the party with the power to take legal action in the event of nonperformance. Since these three interpretations often diverge in practice, for example in cases of a third-party beneficiary under a contract, any adequate analysis of a claim must decide between them or find another explanation of the relational aspect of a claim. Moreover, when one is told that "a claim just is a constraint on behavior" (202), one would like some clear account of the nature of a moral constraint. In a book on moral rights, Thomson need not decide whether the constraint of legal duties consists merely in the attached penalty for violation, but she can hardly ignore the question of whether moral constraint consists entirely or in part in moral sanctions, especially when she concludes that "there do not appear to be any simple, nontrivial theses connecting claims with the permissibility of enforcing them" (116).

If Thomson's analysis of a claim were adequate, would it, as she alleges (200), suffice to explain what all rights are? I believe not. Even supposing that every power were some sort of ability to extinguish or generate a claim, one needs to explain what distinguishes a specifically moral ability from a nonmoral one. I have the ability by persistent pleading to cause my wife to promise to prepare my favorite dish for dinner, but it is she, and not I, who has the ***moral*** power to confer upon me any claim that she do so. What is needed here is some moral analogue of Hohfeld's legally operative facts, but how might one explain how certain facts could be morally operative?

And even if every part of a cluster right were definable in terms of claims, there remain problems in explaining the unity of that right.

> Many of the most familiar rights . . . are clusters of rights, picked out and marked off by mention of a positive value, the rights in the cluster being those that in one or another way 'protect' the value (285).

This is plausible in the case of the right to liberty, but does *the value* of liberty define the content of this right? This seems to imply that one's right does not extend to valueless liberties, much less liberties that it would be harmful to exercise on this occasion. And is a trivial value sufficient to unify a cluster right? Thomson suggests that the value of the green cake with the walnut on top is what unifies her right to that cake (287–8). But surely it cannot be its value to her which confers on her any right to it. Perhaps she made it herself or the hostess promised it to her. If so, it would be the value of making or promise keeping which is central here, not the value of that tiny cake to which she has a cluster right. Moreover, precisely how do the Hohfeldian elements protect the unifying value? Thomson acknowledges by her use of quotation marks that something like Neil MacCormick's conception of normative protection is called for here, but she offers no account of what specifically moral protection might be.

Thomson asks what feature of us is the source of our natural claims. She surveys two alternative possibilities.

> (i) We would have had no claims if we had not been subject to moral law, for if we were not subject to moral law, then *a fortiori* morality would not require of us that we accord any claims. . . . (ii) We would have had no claims if we had been subject to moral law, but a moral law that requires what is incompatible with *Y*'s being, in respect of *X*, under the behavioral constraint that *X*'s having a claim against *Y* consists in (222).

What is puzzling about both alternatives is that they are primarily features of the duty bearer rather than the right holder. This calls into question her presumption that it is something about *us* that explains why *we* have claims.

She goes on to argue (215–21) that (ii) would have been true of us if and only if we had had no interests that were not identical with the interests of the group as a whole. Hence, one source of our natural claims is our inherently individual interests. But why imagine that the content of the moral law that applies to us is determined by the nature of our interests? Thomson seems to presuppose some sort of teleological theory of obligation here. One wonders how she can also assert "There is more to morality than value: there are also claims" (148). No doubt her distinction between value and good for a person (150–2) is relevant here. What remains in doubt is how, if at all, this distinction can be incorporated into a coherent and plausible view of moral obligation.

What basic rights do we have? Apart from our cluster right to liberty and various powers and immunities, we have only pure social claims arising from someone giving her word or from property law and natural claims that others not commit trespass against us or cause us harm or non-belief-mediated distress. In Part II of *The Realm of Rights,* Thomson appeals to our presystematic moral judgments to show that we do have these rights. She defends her thesis that we have no additional claims by examining plausible counterinstances and suggesting that each has morally unacceptable implications. Although I suspect that her "limits thesis" (273–4) is mistaken, her argument for that thesis is entirely appropriate, and she is surely correct to conclude that the burden of proof lies with anyone who would assert that we have any additional moral claims.

The Realm of Rights does not pretend to be a complete and systematic theory; what it proposes is a way of looking at rights.

> To have a right is to have a kind of moral status, so working out what a right is comes to the same as working out what people ought or ought not to do, may or may not do, given a person has a right (373).

Instead of beginning with some normatively neutral definition of "a right" and then going on to investigate what rights we have, Thomson appeals to the moral implications of rights throughout. The result is a book that is immensely rich in moral insight and illuminates, as too few philosophical treatises do, what is of practical importance in rights. But a moral theory contains explanatory theses as well as object-level moral judgments (30). On this level, especially regarding the sources of our rights, this book leaves many questions unanswered. Fortunately, there is hope that Thomson's view of rights can be helpful as she, and we, continue to probe these questions.

CARL WELLMAN

Washington University

NOTES AND NEWS

The editors regret the untimely death of Professor Gerald Feinberg on April 21, at the age of 58. Professor Feinberg obtained his bachelor's degree in 1953 and his doctorate in physics in 1957, under T. D. Lee. Both degrees were from Columbia. He had been a faculty member of the Columbia Physics Department since 1958. Professor Feinberg is best known for his contributions to the theoretical physics of neutrinos and of interatomic forces, and for the theory of tachyons, hypothetical relativistic particles that travel faster than light. He was the author of numerous books and articles on a wide variety of subjects, including the philosophical foundations of physics. His work continued until the very end of his life. His last philosophical paper will appear in these pages in the next few months.

The Planning Committee for the XXth World Congress of Philosophy invites suggestions for the overall theme of the Congress, which will be held in Boston on August 10–16, 1998. The theme must be specified in the Planning Committee's formal proposal to host the World Congress, which will be considered by the General Assembly at the XIXth World Congress in Moscow in August 1993. It should be a topic of broad interest to philosophers in the United States and abroad. Suggestions should be sent to the Chair of the Planning Committee, Richard De George, Phil. Dept., Univ. of Kansas, Lawrence, KS 66045. Suggestions regarding potential sources of financial support from governmental, foundation, and corporate sources are also invited.

Professor and the late Mrs. Colin Turbayne established an annual Berkeley Essay Prize competition in conjunction with the Philosophy Department at the University of Rochester. They are happy to announce that the prize winner in the third year is Lisa Dowling from Yale University. The next deadline for submitting papers is November 1. Submissions on any aspect of Berkeley's philosophy are welcome. Essays should be unpublished, written in English, and less than 5,000 words. Submissions should be sent to Richard Feldman, Chair, Phil. Dept , Univ. of Rochester, Rochester, NY 14627-0078.

The Editorial Board of *The Philosophical Quarterly* would like to announce the topic of its 1992 essay competition: *Scholasticism: Old and New*. Contributions may be on any topic falling within this general theme; especially welcome, however, will be papers which explore connections between issues drawn from different historical periods, and in which areas of current philosophical interest are discussed. The final date for submission is November 1. Essays should not be longer than 6,000 words and should conform to the usual stylistic requirements. Two copies of each essay are required; all entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and nonwinning entries judged to be of sufficient quality will be published. For further information, you may write to Executive Editor, *The Philosophical Quarterly*, Univ. of St. Andrews, Scotland KY16 9AL.

NEW FROM BLACKWELL

LAST WRITINGS ON THE PHILOSOPHY OF PSYCHOLOGY
Volume 2: The Inner and the Outer, 1949 - 1951
LUDWIG WITTGENSTEIN
Edited by GEORG HENRIK VON WRIGHT & HEIKKI NYMAN
Translated by C.G. LUCKHARDT & M.A.E. AUE
January 240 pages
0-631-18256-X cloth $44.95

HEIDEGGER
A Critical Reader
Edited by HUBERT DREYFUSS & HARRISON HALL
June 288 pages
0-631-16342-5, paper $19.95
0-631-16341-7, cloth $44.95

DERRIDA
A Critical Reader
Edited by DAVID WOOD
May 256 pages
0-631-16121-X, paper $19.95
0-631-16102-3, cloth $54.95

PRAGMATIST AESTHETICS
Living Beauty, Rethinking Art
RICHARD SHUSTERMAN
January 288 pages
0-631-18236-5, paper $21.95
0-631-16445-6, cloth $49.95

PHILOSOPHICAL AESTHETICS
An Introduction
Edited by OSWALD HANFLING
February 496 pages
0-631-18035-4, paper $19.95
0-631-18034-6, cloth $54.95

HOLISM
A Shopper's Guide
JERRY FODOR & ERNEST LEPORE
January 272 pages
0-631-18193-8, paper $14.95
0-631-18192-X, cloth $36.95

LANGUAGE FOR HUMANS AND ROBOTS
PATRICK SUPPES
January 384 pages
0-631-18262-4, cloth $44.95

A ROUSSEAU DICTIONARY
N.J.H. DENT
The Blackwell Philosopher Dictionaries
June 260 pages
0-631-17569-5, paper $19.95
0-631-17568-7, cloth $49.95

UNNATURAL DOUBTS
Epistemological Realism and the Basis of Scepticism
MICHAEL WILLIAMS
Philosophical Theory Series
February 384 pages
0-631-16251-8, cloth $44.95

THE CONCEPT OF TIME
MARTIN HEIDEGGER
Translated by WILLIAM McNEILL
February
0-631-18425-2, paper $14.95
0-631-18216-0, cloth $34.95

BELIEF AND MEANING
AKEEL BILGRAMI
March 320 pages
0-631-17776-0, cloth $54.95

A HEGEL DICTIONARY
MICHAEL INWOOD
The Blackwell Philosopher Dictionaries
June 260 pages
0-631-17533-4, paper $19.95
0-631-17532-6, cloth $49.95

NOÛS
Quarterly, ISSN 0029-4624
Subscription Rates Volume 26 1992:
Institutional: $66.00 (N. Am.) $77.00 (R.O.W.)
Individual: $32.00 (N. Am.) $43.00 (R.O.W.)
(Canadian residents please add 7% GST)

BLACKWELL
Publishers

Three Cambridge Center, Cambridge, MA 02142
CALL TOLL-FREE (800) 445-6638
Vermont & Canada dial (802) 878-0315 (MasterCard/Visa/Amex)

CHICAGO

AMERICA'S PHILOSOPHICAL VISION

John E. Smith

In these previously uncollected essays, Smith argues that American philosophers like Peirce, James, Royce, and Dewey have forged a unique philosophical tradition—one that is rich and complex enough to represent a genuine alternative to the analytic, phenomenological, and hermeneutical traditions which have originated in Britain or Europe.

Paper $15.95 224 pages
Library cloth edition $39.95

THE DIALECTICAL NECESSITY OF MORALITY

An Analysis and Defense of Alan Gewirth's Argument to the Principle of Generic Consistency

Deryck Beyleveld

"A scholarly and detailed study of Alan Gewirth's moral philosophy in the light of the criticisms that a number of us have offered of it. It would be hard to imagine a more exhaustive treatment of its subject matter. Beyleveld's judgments seem to me to be illuminating for the most part and largely justified."
—W. D. Hudson

Paper $22.50 562 pages
Library cloth edition $55.00

DESCARTES' METAPHYSICAL PHYSICS

Daniel Garber

Garber's novel and controversial book is the the first to examine Descartes' physics and his philosophy not as separate enterprises but as what they were—a systematic attempt to deal with *all* knowledge.

Paper $23.95 408 pages
Library cloth edition $60.00

Science and Its Conceptual Foundations series

Now in Paper

ESSAYS ON INDIVIDUALISM

Modern Ideology in Anthropological Perspective

Louis Dumont

"As with every learned, perceptive mind, we have much to learn from Dumont's bold forays into modern ideology."—J. G. Merquior, *Critical Review*

Paper $14.95 296 pages

THE FOUCAULT EFFECT

Studies in Governmentality

Edited by Graham Burchell, Colin Gordon, and Peter Miller

In this long-awaited sequel to *Power/Knowledge* Foucault examines the activity of government both in its present form and within a historical perspective. Two lectures by and an interview with Foucault are supplemented here with essays on the topic of governmentality by internationally renowned scholars.

Paper $19.95 320 pages
Library cloth edition $58.00

PHENOMENOLOGY AND DECONSTRUCTION

VOLUME I

The Dream is Over

Robert Denoon Cumming

"The result of Cumming's approach is that one gets the sense of the organic development of phenomenology synthetically as a complex of interactions, a living totality. Written with remarkable clarity and grace, *The Dream is Over* is required reading for anybody seriously interested in intellectual history."
—Hazel E. Barnes

Paper $17.95 272 pages
Library cloth edition $49.95

Distributed for the Center for the Study of Language and Information

THE LANGUAGE OF FIRST-ORDER LOGIC

SECOND EDITION

Including the Macintosh™ program Tarski's World

Jon Barwise and John Etchemendy

"The emphasis throughout is on the user *doing* logic by working the many problems in the text. Recommended unreservedly as a fascinating approach to a fascinating subject."—*Choice*, reviewing the first edition

Paper $34.95 318 pages
Includes 3.5 inch diskette

SITUATION THEORY AND ITS APPLICATIONS

VOLUME 2

Edited by Jon Barwise, Jean Mark Gawron, Gordon Plotkin, and Syun Tutiya

Essay topics include natural language semantics, philosophical issues concerning information, mathematical applications, and the visual representation of information in computer systems

Paper $26.95 638 pages
Library cloth edition $59.95

THE UNIVERSITY OF CHICAGO PRESS
5801 S. ELLIS AVE., CHICAGO, IL 60637

CAMBRIDGE PRESENTS...

Nietzsche's Philosophy of Art

Julian Young

Julian Young argues that Nietzsche's thought about art can only be understood in the context of his wider philosophy. He discusses the dramatic changes in Nietzschean aesthetics against the background of the celebrated themes of the death of God, eternal recurrence, and the idea of the *Übermensch*.

41124-6 Hardcover $44.95

The Aesthetic as the Science of Expression and of the Linguistic in General

Benedetto Croce

Colin Lyas, Translator

"The case for a new translation of Croce's *Estetica* is overwhelming...for all its influence it has never been widely read in the English-speaking world...this new translation is fully set to put right this anomaly."

—*Richard Wollheim*

35216-9 Hardcover $49.95

35996-1 Paper $16.95

The Language of Art History

Salim Kemal and

Ivan Gaskell, *Editors*

This volume offers a range of responses to some crucial issues generated by the relationship between the art object and language in art history.

Contributors: *Michael Baxandall, Ivan Gaskell, Andrew Harrison, Carl R. Hausman, Salim Kemal, Catherine Lord, Jose Bernadete, Jean-Francois Lyotard, Stanley Rosen, Richard Shiff, David Summers, Richard Wollheim.*

Cambridge Studies in Philosophy and the Arts 1

35384-X Hardcover $54.95

On the Nature and Existence of God

Richard M. Gale

"...a genuine contribution to the literature in the philosophy of religion." —*Nelson Pike, University of California, Irvine*

40300-6 Hardcover $44.50

Hegel's Political Theology

Andrew Shanks

This study attempts to grasp the continuing contemporary relevance of Hegel's political theology, which the author interprets as a uniquely radical critique of every sort of religious authoritarianism.

40321-9 Hardcover $49.50

Now in paperback...

Eberhard Jüngel

An Introduction to his Theology

J. B. Webster

As a comprehensive survey of Jüngel's work as New Testament scholar, systematic theologian and philosopher, this volume focuses on his discussion of theological language, the role of Christology, and the doctrines of God and man.

30708-2 Hardcover $54.95

42391-0 Paper $16.95

The Trespass of the Sign

Deconstruction, Theology and Philosophy

Kevin Hart

Jacques Derrida's theory of deconstruction has been depicted either as a direct attack against philosophy or a displaced negative theology. According to this innovative study, neither view is correct.

35481-1 Hardcover $54.95

42382-1 Paper $16.95

PHILOSOPHICAL TOPICS

Archaeological Typology and Practical Reality
A Dialectical Approach to Artifact Classification and Sorting
William Y. Adams and Ernest W. Adams
39334-5 Hardcover $64.50

Philosophy and Biblical Interpretation
A Study in Nineteenth Century Conflict
Peter Addinall
40423-1 Hardcover $54.50

Alban Berg
Master of the Smallest Link
Theodor W. Adorno
Christopher Hailey and Juliane Brand, Translators
33016-5 Hardcover $39.95

The Meanings of Death
John Bowker
39117-2 Hardcover $27.95

The Subject of Modernity
Anthony J. Cascardi
41287-0 Hardcover $49 95
42378-3 Paper $16 95

Film and Phenomenology
Towards a Realist Theory of Cinematic Representation
Allan Casebier
41132-7 Hardcover $49 95

The Ant and the Peacock
Altruism and Sexual Selection from Darwin to Today
Helena Cronin
Foreword by John Maynard Smith
32937-X Hardcover $39.50

Logic and Information
K. Devlin
41030-4 Hardcover $32 95

Myth, Truth and Literature
Towards a True Post-Modernism
Colin Falck
36256-3 Hardcover $44 95
42384-8 Paper $15.95

Goethe's History of Science
Karl J. Fink
40211-5 Hardcover $47 95

Plausible Worlds
Geoffrey Hawthorn
40359-6 Hardcover $44 50

Mortal Questions
Thomas Nagel
40676-5 Paper $8 95

A World Theology
The Central Spiritual Reality of Humankind
N. Ross Reat and Edmund F. Perry
33159-5 Hardcover $49 95

Arguments of Augustan Wit
John Sitter
41120-3 Hardcover $44 95

Eminent Economists
Their Life Philosophies
Michael Szenberg
38212-2 Hardcover $34 95

Return to Freud
Jacques Lacan's Dislocation of Psychoanalysis
Samuel Weber
Michael Levine, Translator
37410-3 Hardcover $47 50
37770-6 Paper $13 95

CAMBRIDGE
UNIVERSITY PRESS

40 West 20th Street, New York, NY 10011-4211 Call toll-free 800-872-7423
MasterCard/VISA accepted Prices subject to change

Princeton

A Prosentential Theory of Truth

Dorothy Grover

In a number of influential articles published since 1972, Dorothy Grover has developed the prosentential theory of truth. Brought together and published with a new introduction, these essays are even more impressive as a group than they were as single contributions to philosophy and linguistics.

Cloth: $39.50 ISBN 0-691-07399-6 *Due June 1992*

New paperback edition, with a new afterword by the author

Women in Western Political Thought

Susan Moller Okin

"Okin has written an engaging, serious, careful, and important work that raises the issues of women and politics in their most elemental and pertinent form. . . A pioneering book." —*Benjamin R. Barber, The New Republic*

"Susan Moller Okin's book lucidly establishes [that] we cannot simply read the great political philosophers as though 'mankind' means all of us."—*Nannerl Keohane, Women and Politics*

Paper: $14.95 ISBN 0-691-02191-0

New in paperback

On Social Facts

Margaret Gilbert

Are social groups real in any sense that is independent of the thoughts, actions, and beliefs of the individuals making up the group? Using methods of philosophy to examine such longstanding sociological questions, Margaret Gilbert gives a general characterization of the core phenomena at issue in the domain of human social life. In her analyses Gilbert discusses the work of such thinkers as Emile Durkheim, Georg Simmel, Max Weber, and David Lewis.

Now in paper $19.95 ISBN 0-691-02080-9

Princeton University Press

41 WILLIAM ST • PRINCETON, NJ 08540
ORDERS 800-777-4726 • OR FROM YOUR LOCAL BOOKSTORE

THE JOURNAL OF PHILOSOPHY

is pleased to announce the publication of its

TWENTY-FIVE YEAR INDEX, 1964–1988

VOLUME LXI TO VOLUME LXXXV

AUTHORS, SUBJECTS, & REVIEWS

Supplement to its

FIFTY-YEAR INDEX, 1904–1953

AND

TEN-YEAR SUPPLEMENT, 1954–1963

$25.00	Cloth—Libraries
20.00	Cloth—Individuals
10.00	Paper—Individuals
8.00	Paper—Individuals (with new subscriptions)

Please send a check or money order to:

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through May 31, 1993. Orders and payment must be received by then. Paper copies will be mailed immediately, but cloth copies will take 6–8 weeks for delivery.

25 NOV 1992

Princeton

Entailment

The Logic of Relevance and Necessity

Volume II

Alan Ross Anderson, Nuel D. Belnap, Jr., and J. Michael Dunn

In spite of a powerful tradition, more than two thousand years old, that in a valid argument the premises must be relevant to the conclusion, twentieth-century logicians neglected the concept of relevance until the publication of Volume I of this monumental work. Since that time relevance logic has achieved an important place in the field of philosophy. Volume II of *Entailment* brings to a conclusion a powerful and authoritative presentation of the subject by some of the foremost scholars working in the area.

Cloth: $95.00 ISBN 0-691-07339-2

Discovering Complexity

Decomposition and Localization as Strategies in Scientific Research

William Bechtel and Robert C. Richardson

This study of the dynamics of theory development focuses on two psychological heuristics, decomposition and localization. William Bechtel and Robert Richardson identify a number of "choice-points" that scientists confront in developing mechanistic explanations and describe how different choices result in divergent explanatory models.

Cloth: $39.50 ISBN 0-691-08762-8

Invitation to Mathematics

Konrad Jacobs

Based on a well-received course designed for philosophy students, this book is an informal introduction to mathematical thinking. The work will be rewarding for all philosophers concerned with mathematical questions. Konrad Jacobs discusses an unusually wide range of topics, including such items of contemporary interest as knot theory, optimization theory, and dynamical systems.

30 halftones. 90 line illustrations

Paper: $29.95 ISBN 0-691-02528-2

Cloth: $80.00 ISBN 0-691-08567-6

Princeton University Press

41 WILLIAM ST PRINCETON, NJ 08540 • ORDERS: 800-777-4726 OR FROM YOUR LOCAL BOOKSTORE

Princeton

Essays on Aristotle's *Poetics*

Edited by Amelie O. Rorty

Aimed at deepening our understanding of the *Poetics*, this collection places Aristotle's analysis of tragedy in its larger philosophical context. In these twenty-one essays, philosophers and classicists explore the corpus of Aristotle's work in order to link the *Poetics* to the rest of his views on psychology and on history, ethics, and politics. The essays address such topics as catharsis, pity and fear, pleasure, character and the unity of action, and the modality of dramatic action.

Paper: $19.95 ISBN 0-691-01498-1
Cloth: $69.50 ISBN 0-691-06872-0

Principles and Proofs

Aristotle's Theory of Demonstrative Science

Richard D. McKirahan, Jr.

By a thorough study of the *Posterior Analytics* and related Aristotelian texts, Richard McKirahan reconstructs Aristotle's theory of *episteme*—science. The *Posterior Analytics* contains the first extensive treatment of the nature and structure of science in the history of philosophy, and McKirahan's aim is to interpret it sympathetically, following the lead of the text, rather than imposing contemporary frameworks on it. In addition to treating the theory as a whole, the author uses textual and philological as well as philosophical material to interpret many important but difficult individual passages.

Cloth: $47.50 ISBN 0-691-07363-5

The Lockean Theory of Rights

A. John Simmons

John Locke's political theory has been the subject of many detailed treatments by philosophers and political scientists. But *The Lockean Theory of Rights* is the first systematic, full-length study of Locke's theory of rights and of its potential for making genuine contributions to contemporary debates about rights and their place in political philosophy. Given that the rights of persons are the central moral concept at work in Locke's and Lockean political philosophy, such a study is long overdue.

Studies in Moral, Political, and Legal Philosophy
Marshall Cohen, Editor
Cloth: 39.50 ISBN 0-691-08630-3

Classics in Philosophy

GREEK PHILOSOPHY
Thales to Aristotle
Third Edition, Revised and Expanded

Reginald E. Allen
1991 0-02-900495-0 $14.95 paper

POLITICAL IDEAS OF ST. THOMAS AQUINAS
St. Thomas Aquinas
0-02-840380-0 $10.95 paper

EIGHTEENTH-CENTURY PHILOSOPHY
Lewis W. Beck, Editor
0-02-902160-6 $12.95 paper

A MODERN INTRODUCTION TO PHILOSOPHY, 3rd Edition
Paul Edwards & Arthur Pap, Editors
0-02-909200-0 $27.95

A SHORT HISTORY OF CHINESE PHILOSOPHY
Fung Yu-Lan
0-02-910980-9 $14.95 paper

NINETEENTH-CENTURY PHILOSOPHY
Patrick Gardiner, Editor
0-02-911220-6 $12.95 paper

INTRODUCTORY READINGS IN AESTHETICS
John Hospers
0-02-915260-7 $14.95 paper

MORAL AND POLITICAL PHILOSOPHY
David Hume
0-02-846170-3 $12.95 paper

DIALOGUES CONCERNING NATURAL RELIGION
David Hume
0-02-846180-0 $10.95 paper

PHILOSOPHY OF THE SIXTEENTH AND SEVENTEENTH CENTURIES
Richard H. Popkin
0-02-925490-6 $12.95 paper

ESSAYS IN PRAGMATISM
William James
0-02-847140-7 $9.95 paper

CRITIQUE OF JUDGEMENT
Immanuel Kant
0-02-847300-3 $12.95 paper

PHILOSOPHY AND TECHNOLOGY
Carl Mitcham & Robert Mackey, Editors
0-02-921430-0 $14.95 paper

THE SPIRIT OF THE LAWS
Baron de Montesquieu
0-02-849270-6 $13.95 paper

NEWTON'S PHILOSOPHY OF NATURE
Isaac Newton
0-02-849700-7 $12.95 paper

A CRITICAL HISTORY OF WESTERN PHILOSOPHY
D.J. O'Connor
0-02-923840-4 $14.95 paper

GREEK AND ROMAN PHILOSOPHY AFTER ARISTOTLE
Jason L. Saunders
0-02-927730-2 $14.95 paper

ETHICS
Benedict de Spinoza
0-02-852650-3 $10.95 paper

TWENTIETH CENTURY PHILOSOPHY
The Analytic Tradition
Morris Weitz
0-02-934990-7 $12.95 paper

MEDIEVAL PHILOSOPHY
John F. Wippel & Allan B. Wolter
0-02-935650-4 $12.95 paper

Attention College Professors: Examination copies will be sent on a 30-day approval basis only. Simply place your request on university stationery, indicating book requested, course title, projected enrollment, and date your next text decision is due. Mail to.

Fp **THE FREE PRESS**
A Division of Macmillan/A Maxwell Macmillan Company
866 Third Avenue, New York, NY 10022

For the best in Philosophy turn to Cambridge

A Structuralist Theory of Logic

Arnold Koslow

"In short, this volume has the makings of a classic. I have rarely found myself so enthusiastic about a recent book in logic. I recommend it in the strongest terms."

—*Isaac Levi, Columbia University*

41267-6 Hardcover $69.50

Probability and the Art of Judgment

Richard Jeffrey

Spanning a period of 35 years, this collection of essays includes some of the classic works of one of the most distinguished and influential philosophers working in the field of decision theory and the theory of knowledge.

Cambridge Studies in Probability, Induction and Decision Theory

39459-7 Hardcover $49.95
39770-7 Paper $16.95

Memorial Essays

A. J. Ayer

A. Phillips Griffiths, Editor

This collection of essays explores the thought of one of Britain's most distinguished philosophers and reflects on his place in the history of philosophy.

Contributors: *Frederick Copleston, Donald Davidson, Ted Honderich, Stephan Korner, Hugh Mellor, Hilary Putnam, Lord Quinton, Stewart Sutherland, John Watling, David Wiggins, Bernard Williams, Richard Wollheim.*

Royal Institute of Philosophy Supplements 30

42246-9 Paper $18 95

Picoeconomics

The Interaction of Successive Motivational States within the Person

George Ainslie

This book examines an elementary human paradox: that we are endangered by our own wishes. Dr. Ainslie uses the results of behavioral experiments to show that there is a basic tendency for both lower animals and human beings to form preferences for the poorer but earlier of two goals, when the poorer goal is close at hand.

Studies in Rationality and Social Change

26093-0 Hardcover $64 95

The Contents of Experience

Essays on Perception

Tim Crane, Editor

This volume represents the best of the latest research on perception, with contributions from some of the leading philosophers in the area. As well as discussing traditional problems, the essays also approach the topic in light of recent research on mental content and representation.

Contributors: *Thomas Baldwin, Tim Crane, E. J. Lowe, Brian O'Shaugnessy, Christopher Peacocke, Paul Snowdon, Michael Tye, J. J. Valberg.*

41727-9 Hardcover $49 95

Available in bookstores or write:

CAMBRIDGE
UNIVERSITY PRESS

40 West 20th St., NY, NY 10011-4211 Call toll-free 800-872-7423
MasterCard/VISA accepted. Prices subject to change.

NEW FROM

PARAGON HOUSE

90 FIFTH AVENUE • NEW YORK NY 10011 • 1-800-PARAGON

COMPLETE CATALOG AVAILABLE.

ETHICAL ARGUMENT

Critical Thinking in Ethics

HUGH MERCER CURTLER

An introduction to the resolution of ethical conflict by means of standard methods of critical thinking. Includes cultural and ethical relativism; three traditional principles of ethical decision making; and the justification of ethical decisions in contrast to rationalization and explanation. Real-life examples enhance the relevance of the material.

160 pp. • Paper • ISBN 1-55778-513-9 • $12.95

PHILOSOPHY OF SCIENCE

JAMES H. FETZER

A foundation for inquiry into the nature of science, the history of science, and the relationship between the two. Investigates the aim and methods of empirical science, the importance of methodological commitments, and the importance of interpretations of probability for understanding scientific laws and explanations.

224 pp. • Paper • ISBN 1-55778-481-7 • $16.95

FOUNDATIONS OF

PHILOSOPHY OF SCIENCE

Recent Developments

Edited by JAMES H. FETZER

This companion reader to *Philosophy of Science* is divided to correspond to the eight chapters of that core text, and includes classic and contemporary selections from Rudolph Carnap, Karl Popper, Carl Hempel, Wesley Salmon, and many others.

592 pp. • Paper • ISBN 1-55778-480-9 • $28.95

"...Bataille's books exert a fascination that never ceases."*

ON NIETZSCHE

GEORGES BATAILLE

Georges Bataille interprets Nietzsche by *becoming* Nietzsche, thinking, probing and extending the great philosopher's moral arguments at every turn. "Bataille's entire work deals with the experience of the edge, that is, living at very limits of life, at the extreme..." —**Bloomsbury Review*

250 pp. • Cloth • ISBN 1-55778-325-X • $22.95

Now in paper

FORMS OF DESIRE
Sexual Orientation and the Social Constructionist Controversy
Edited by **Edward Stein**
384 pp $15 95/paper

JE, TU, NOUS
Toward a Culture of Difference
Luce Irigaray
Translated from the French by **Alison Martin**
216 pp $45 00/cloth $13 95/paper

THE NATURE OF MORAL THINKING
Francis Snare
176 pp $49 95/cloth $14 95/paper

NIETZSCHE'S CASE
Philosophy as/and Literature
Bernd Magnus, Stanley Stewart, and **Jean-Pierre Mileur**
320 pp $32 50/cloth $16 95/paper

THE PHILOSOPHY OF THE LIMIT
Drucilla Cornell
224 pp $45 00/cloth $14 95/paper

SITUATING THE SELF
Gender, Community, and Postmodernism in Contemporary Ethics
Seyla Benhabib
200 pp $32 50/cloth $16 95/paper

Philosophy
from **ROUTLEDGE**

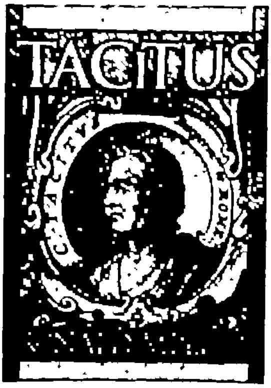
BEING A PHILOSOPHER
The History of a Practice
David W. Hamlyn
224 pp $29 95/cloth

DECONSTRUCTION AND THE POSSIBILITY OF JUSTICE
Edited by **Drucilla Cornell, Michel Rosenfeld,** and **David Gray Carlson**
438 pp $32 50/cloth $16 95/paper

FEMINIST EPISTEMOLOGIES
Edited and with an introduction by
Linda Alcoff and **Elizabeth Potter**
Thinking Gender
256 pp $49 95/cloth $13 95/paper

FIRST PERSON PLURAL
Multiple Personality and the Philosophy of Mind
Stephen E. Braude
288 pp $40 95/cloth

TACITUS
Ronald Mellor
200 pp $25 00/cloth



ROUTLEDGE
Available at better bookstores or from
ROUTLEDGE
29 West 35th Street,
New York, NY 10001-2299

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709-M Philosophy Hall, Columbia University
New York, New York 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 7
JULY 1992

page
331 *Caesar's Wife: On the Moral Significance of Appearing Good* Julia Driver
344 *Semantic Supervenience and Referential Indeterminacy* James Van Cleve
362 *Bayesian Conditionalization Resolves Positivist/Realist Disputes* Jon Dorling

BOOK REVIEWS
383 W. V. Quine and Rudolf Carnap: *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work* Barry Stroud

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027 (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 7, JULY 1992

CAESAR'S WIFE: ON THE MORAL SIGNIFICANCE OF APPEARING GOOD*

SUETONIUS and Plutarch both report that Caesar divorced his wife Pompeia because she was suspected of adultery. In the story related by them, a man named Clodius sneaked into Caesar's house dressed as a woman in order to rendezvous with Caesar's wife, apparently with her complicity. Clodius was apprehended, and Pompeia fell under suspicion. It is not clear that she was actually involved with Clodius. In any case, actual guilt or innocence was irrelevant to Caesar. Caesar's wife must be above suspicion. Since her behavior resembled guilty behavior, she is blamed.[1]

I shall explore the question of whether an action that is not immoral in virtue of any independent reasons can be bad, or immoral, in virtue of resembling an immoral action. The answer will be "yes." What is surprising about this answer is that it means that, in an important class of these cases, the immorality of some actions will depend completely on the mistaken beliefs of other people.

I THE MORAL SIGNIFICANCE OF RESEMBLANCE

That an action's resemblance to an immoral act may be morally significant has been noted before, but with little accompanying analysis. Kant, for example, relies on a resemblance criterion in his *Lectures on Ethics*[2] when discussing duties we have toward animals:

---

* This paper has benefited from conversations with Jonathan Adler, Sarah Buss, John Carroll, Dick Moran, Connie Rosati, Roy Sorensen, and William Williams; from the Philosophy and Public Affairs Society of New York in February 1991, especially Susan Brison, Gertrude Ezorsky, and Frances Kamm; and from the 1991 Pacific Division Meeting of the American Philosophical Association, especially Hugh Lehman and Terrance McConnell

[1] See Plutarch, *The Lives of the Noble Grecians and Romans*, John Dryden and Arthur Hayden Clough, trans. (New York: Random, 1864), pp. 859–60.

[2] Louis Infield, trans (Indianapolis: Hackett, 1963).

0022-362X/92/8907/331–43

since animals are similar to human beings, cruelty toward animals will undermine genuine duties toward other persons. Thus, one ought to refrain from cruelty toward animals in pursuing one's duty toward other *persons*.

> Animal nature has analogies to human nature, and by doing our duties to animals in respect to manifestations which correspond to manifestations of human nature, we indirectly do our duty towards humanity. Thus, if a dog has served his master long and faithfully, his service, on the analogy of human service, deserves reward . . . (*ibid.*, p. 239).

If a person behaves viciously toward an animal he is in danger of corrupting himself by failing to reinforce his benevolent tendencies —tendencies that should be cultivated to support the sense of duty toward humanity. Since animals resemble humans, and animal behavior resembles human behavior, cruelty toward animals may lead people to assimilate this with cruelty toward humans.

Jane English,[3] after describing a Kantian account of why it is wrong to mistreat animals, extends the analysis by citing resemblance as a factor in deciding on the permissibility of abortion.

> . . . the similarity of a fetus to a baby is very significant. A fetus one week before birth is so much like a new born baby in our psychological space that we cannot allow any cavalier treatment of the former while expecting full sympathy and nutritive support for the latter (*ibid.*, p. 175).

English further suggests that resemblance explains why we show respect for dead bodies, even though they are not persons.

No doubt, resemblance exerts a powerful influence over human psychology. According to Hume,[4] it is resemblance between people which helps to explain the force of sympathy:

> . . . this resemblance must very much contribute to make us enter into the sentiments of others and embrace them with facility and pleasure. Accordingly we find, that where, beside the general resemblance of our natures there is any peculiar similarity in our manners, or character, or country, or language, it facilitates the sympathy (*ibid.*, p. 318).

Indeed, it is no accident that during times of war the enemy is often portrayed as being very different from us, since this helps to cut them off from any sympathy we would feel.

---

[3] "Abortion and the Concept of a Person," in *Morality in Practice*, James Sterba, ed. (Belmont, CA: Wadsworth, 1988).

[4] *A Treatise of Human Nature*, L. A. Selby-Bigge, ed. (New York: Oxford, 1967).

This discussion shows that, in addition to leading to self-corruption, resemblance is also relevant because of its effect on others. Someone witnessing a late abortion may grow callous toward newborns. Also, people are liable to make mistakes about what it is that they are seeing. If someone witnesses an action that she *thinks* is directly immoral, then the witness's morals could be corrupted. A person who witnesses the desecration of some bodies, for example, may become callous and brutal.

Broadly speaking, then, there are two ways in which resemblance can be morally significant.

(1) First of all, resemblance causes us to assimilate, psychologically, different types of actions. Because of our "coherence of attitudes," we need to be sensitive to resemblance so that we do not corrupt ourselves or others.[5] Psychologically, we are so constructed as to be influenced by resemblance in the making of our moral judgments. English argues that our attitudes and sympathies, which reinforce our moral principles, need to form a coherent set, thus: "If our moral rules allowed people to treat some person-like nonpersons in ways we do not want people to be treated, this would undermine the system of sympathies and attitudes which makes the ethical system work" (*ibid.*). If one performs actions that look immoral, one may develop a taste for doing the real thing, or a truly unpleasant disposition. So, resemblance can directly cause harm by breaking down resistance to certain actions. By analogy, our sympathies are dampened and our attitudes may shift in a bad direction. A person who sees a 7-month fetus thrown into a garbage pail and an 8-month premature infant in intensive care may become brutalized to the suffering of these infants because the resemblance between fetus and infant will cause her feelings to change. Likewise, some actions will be good in virtue of resembling independently good actions. Not only should we refrain from hurting animals, but maybe we should even be nice to them to encourage similar behavior toward humans. Kant's faithful dog should be rewarded, so that agents also reward faithful persons.

(2) Second, resemblance can lead people to make mistakes—to misconstrue an action's moral character.[6] A person who wears fake

---

[5] See English, pp. 174–5 for a brief discussion of this.

[6] Hume notes (p. 61) that resemblance is a great source of intrapersonal error because it leads to mistakes in reasoning: "there are few mistakes in reasoning; which do not borrow largely from that origin. Resembling ideas are not only related together, but the actions of the mind, which we employ in considering them, are so little different, that we are not able to distinguish them. This last circumstance is of great consequence; and we may in general observe, that when-

fur could be mistaken for a person who wears real fur. If an action resembles an immoral action, then it is not surprising that it could be misunderstood as independently immoral in some contexts. Misconstrual, in turn, has its own corrupting influences. Such misconstrual can corrupt the morals of others, in addition to causing other harms. This is the most interesting claim regarding the moral significance of resemblance, because it means that the immorality of some actions depends on the agent realizing that their true nature will be misunderstood.

The crucial difference between the two ways in which resemblance is significant is that the second involves a misconstrual of the action. In the first case, the person performing the action or witnessing it may know full well the difference between, for example, the dead person and the live person. It is just that the resemblance between them is forceful enough that perhaps the abuse of the former would lead to decreased respect for the latter. In the second class of cases, however, the witness to the action mistakes it for some other action. These mistakes or misconstruals can lead to unfortunate consequences in one of two ways. First, a person can witness an action and mistake it for another action that is immoral. The witness does not believe that the immoral act is immoral, and this leads to the witness' doing or condoning that type of action. For example, in the Caesar's wife case, the witness thinks that Pompeia has commited adultery when in fact she has not, and comes to the conclusion that adultery must be all right after all. Second, a person can witness an action and mistake it for another action that he does believe to be immoral. Still, witnessing the supposedly immoral act causes there to be less pressure on him to avoid those immoral acts. Thus he will go ahead and do them. Thus the witness might see Pompeia leave the door open for Clodius at midnight, think that she is planning an adulterous liason when in fact she is not, but, knowing that adultery is immoral, do it anyway. If she can get away with it, I can too. This is the more plausible sort of scenario in the Caesar's wife case.

Also, misconstrual could lead to good consequences, if an action is mistaken for a good one because of its resemblance to an independently good action. Consider George Bush donating blood. This donation of blood looks like a good action because donating blood

---

ever the actions of the mind in forming any two ideas are the same or resembling we are very apt to confound these ideas, and take the one for the other." The issue of concern in this paper is that of interpersonal error that is the result of resemblance.

helps others. There is a sense in which George Bush's donating blood is a waste, however. Given all the secret service people who need to guard him to and from the bloodmobile, his donation will cost the tax payers a lot of money. His action, on scrutiny, is dissimilar to the ordinary citizen's. Still, it resembles an ordinary person's sacrifice, and is easily mistaken for such, and thus it has great symbolic value. This symbolic value can be cashed in in the hard utilitarian currency of consequences. It will motivate others to give blood —those who can give blood efficiently.

These observations give us a place to start in formulating a principle of action.

> (CP) An action is itself immoral (moral) if it resembles an independently immoral (moral) act.

An independently immoral act is an act that is immoral for reasons other than resemblance to something immoral. Actions immoral due to their resemblance to independently immoral actions I call ***mimetic immoralities***. Both mimetically immoral actions and independently immoral ones are immoral, but they are different descriptively, and also in terms of the degree of condemnation typically associated with them.

(CP) is false. It is open to fairly obvious objections and counterexamples. First of all, a slippery slope is easily generated. It is immoral to drive 100 mph in a school zone. And if it is immoral to drive n mph in a school zone, it is also immoral to drive $n-1$ mph in a school zone (since the speeds are nearly indistinguishable). Thus, it is immoral to drive 10 mph in a school zone. The second step of this induction would be supported by (CP) as it stands at this point. Further, some actions that resemble the genuinely immoral are permissible and even admirable. An actress may innocently perform a pretend murder; lifeguards are applauded for administering mouth-to-mouth recusitation. Nevertheless, (CP) has a grain of truth. As above cases illustrate, there is something about resemblance which is morally significant. One of my goals here is to articulate that significance.

II AVOIDING MISCONSTRUAL

Most of us at some time or other are faced with the choice of whether to do something that *could* be misconstrued as immoral by others. A person who believes that killing animals for fur is immoral, and thus will not wear a real fur coat, may still feel uneasy about wearing a good imitation; or a person who runs away from the scene of a crime because he needs to make it to the bank on time may worry that his action could be misconstrued as the action of a criminal; or the feminist who goes into a strip-tease joint just to use the toilet might have a similar worry that his action could be misconstrued. How can we best explain the moral unease?[7]

Intuitions seem inconveniently split: on the one hand, one is tempted to say: "Who cares about what others think? I'm going to do what I think is appropriate, what *I* know to be perfectly okay." This expresses moral individualism, a strain of thinking that runs through our common-sense ethics. It pictures moral obligations as autonomous relations, for example. Moral individualism can be argued for a number of ways. One approach is the idea that moral obligations can only be voluntarily acquired, that we have ultimate control over the moral aspects of our lives. If we accept the moral relevance of *appearances*, then morality falls into the hands of the illogical and the ill-informed. What is unpleasant here is that, if one is prohibited morally from performing such actions, another bit of one's autonomy is lost. There is nothing at all intrinsically wrong with these actions, yet, sensitive to consequences, the agent must refrain from doing them. And the consequences are brought about because others misunderstand the agent's action. That is particularly galling. Why should the agent be responsible for *that?*

The moral individualist need not shirk responsibilities. Indeed, he frequently stresses that an agent is responsible for his own actions. He maintains, however, that how others merely *perceive* actions is irrelevant. This theme breaks to the surface in some of Bernard Williams's[8] writing on integrity. Williams complains that utilitarianism, for example, fails to give adequate weight to the concerns and projects of the individual. In his critique of utilitarianism he has pointed out (in this vein) that, if utilitarianism is correct, then some actions that are prima facie wrong are actually right; and that this is psychologically too jarring, and demands too much of the moral agent. For example, the utilitarian would hold that, if killing another human being would clearly maximize good consequences, then the agent should kill. Otherwise, she is responsible for whatever bad consequences are the result of the failure to act. Thus, an action like killing, which seems terribly wrong, could be required—on the utilitarian view, it could turn out to be the right thing to do. Thus, the agent may be required to do something that goes against his most deeply held convictions and commitments. This, Williams argues, violates the agent's integrity.

---

[7] Some might maintain that the unease is caused by a fear that one's reputation will suffer from the misconstrual While I think this is a genuine concern—no one wants to be thought of as immoral—it does not exhaustively explain the unease. One is worried that there is something wrong in the action, even though one's motives may be pure.

[8] "A Critique of Utilitarianism," in *Utilitarianism: For and Against,* J. J. C. Smart and Williams (New York: Cambridge, 1973), pp. 77ff.

The other side of that coin is that some actions that are prima facie permissible, or even right, are really wrong—e.g., mimetically immoral acts. A utilitarian would require of the moral agent that he not wear the imitation fur, if the action is open to misconstrual and would corrupt or deeply hurt others.

The moral individualist will not appreciate having her actions constrained by the false beliefs of others. This is because it entails that there is some aspect of her moral life over which she has very little control. It makes the quality of her actions depend on the interpretation of others. On the other hand, in contrast to moral individualism, one does not want to give a false impression, especially when it could influence other people. If one's behavior is influential, then one should not even give the appearance of improper behavior.[9]

One explanation for the muddled intuitions is that we often seem to buy into a moral safety argument in trying to figure out how we ought, or ought not, act. If I cannot decide whether an act is permissible or not, I shall play it safe (morally) and refrain from doing it. For example, if a person feels that killing animals is wrong because they are sentient, she may be uncertain about where to draw the line. Is it all right to eat shell fish? To play it safe, she may just refuse to eat any animal at all.

In Amish societies, this tendency to "play it safe" has often led to splits in the group. Suppose that there is a prohibition against wearing more than three buttons on one's clothing at any given time. This might be because buttons are viewed as personal adornment, and thus frivolous. Some people, in order to make absolutely sure that they are not behaving frivolously, will begin to wear even fewer buttons. This tendency seems to change standards of appropriateness—because one does not want to even give the appearance of frivolousness one will wear only 2 buttons, or only 1, or even no buttons at all. The desire here is to play it safe. But in doing this eventually what was once optional goodness becomes required.

Perhaps this is so with moral appearances as well. Perhaps we are uncertain about the moral nature of these actions because some of their features do seem to make them immoral. The uncertainty is caused by using an analogy with the independently immoral acts, and concluding that these actions are immoral, too. To make absolutely sure we are moral people, then, we might not perform mimetically immoral actions.

---

[9] It is also possible that this type of misconstrual will have good effects. Perhaps, for example, these actions encourage people to be more charitable since once in a while they will discover that they misinterpreted the nature of an action they witness. So, if the misconstrual occurs, *and is then discovered,* it may actually do some good as well.

Yet I do not think that moral safety considerations can fully explain the phenomenon. Consider again the fur-coat case: here, the agent knows that the coat was manufactured in an acceptable manner: no leg-hold traps are used in the manufacture of artificial fur. No fur trader received a benefit when the agent obtained the coat. There is no uncertainty in this case about whether wearing the coat can be described as independently immoral, because the reasons used to show that real fur is immoral are lacking in this case. The analogy does not exist at this level of reasons. Yet the act still seems tainted morally—and purely in virtue of its resemblance to the independently immoral act of purchasing a real fur coat.

To make the matter a little more clear: in the imitation-fur-coat case, wearing it would not seem bad if one wore it only among other persons fully aquainted with the facts. And in this case it is certainly possible to be fully acquainted with the facts. If the agent were to wear the coat only in the presence of close friends who knew the origins of the coat (or, better yet, in private), moral discomfort would almost completely disappear. Thus, audience relativity is an important factor. An action may be immoral relative to one audience and not to another.

An actor on the stage performs actions that resemble those which others perform off stage and in earnest. So, the actor playing Othello pretends to strangle, and his action resembles a genuine strangling. Yet the context is such that it is clear to everyone in the audience that the actor is not really strangling anybody. In an artificial setting such as this there is no danger, because the audience is not liable to misconstrual. But, in real life, misconstrual of one's actions will almost certainly take place.

For these reasons, the explanation for why such mimetically immoral actions are immoral is that these actions, because they are open to reasonable misconstrual, lead to bad consequences. Whenever we are in a situation where our actions can be so misconstrued, we must be careful, because what we do may set an example for others.

III. SETTING AN EXAMPLE

In his *Lectures on Ethics,* Kant stresses the importance of good examples in moral life: they demonstrate the practical possibility of moral acts. "A bad example, however is a stumbling block, and gives occasion for two evils, its imitation as a pattern and excuse" (*op. cit.*, p. 111). Further, Kant writes:

> It is true that I can scarcely be blamed if another mistakes and misapplies what I have done and it corrupts his morals, although I meant it otherwise and it had no immoral effect on me; but, though I am not liable for his misconstruction, I ought none the less to exercise restraint to avoid giving occasion for any such *scandalum* (*op. cit.*, p. 111).[10]

I take it that Kant is saying that, though one is not in any sense required to avoid performing what I call mimetically immoral acts, one ought not—in some weaker sense—perform them. That is, one is not responsible for the misconstrual—and the example Kant goes on to discuss makes it obvious that one is not responsible even if one can foresee that the misconstrual will take place. Still, one should not perform such actions, because one would be setting a bad example for others.

Extrapolating from Kant's discussion of animals, my guess is that Kant would hold that we ought not to perform mimetically immoral actions because to perform such actions would undermine our moral commitment to avoid doing independently immoral actions. The duty is an "indirect" one. If there is a danger of self-corruption one should not perform these actions. Also, as the above passage indicates, even when there is no danger of self-corruption one should refrain because the mistaken individual who witnesses the action stands in danger of corruption.

Usually these indirect duties are described as pseudo duties one has to animals and inanimate objects in virtue of the duties one has toward rational beings. They are postulated to explain why we think cruelty to animals, and the wanton destruction of an object, are immoral acts—even though the objects are not rational. The real duties, or the direct duties, are those we have toward rational beings as ends in themselves. There is nothing intrinsically wrong with hurting animals, but one should refrain from hurting them because such behavior will lead to a lack of concern for rational beings. Here the analysis depicts the first role of resemblance—that is, the effect it has on our moral thinking in virtue of our tendency to assimilate behaviors that resemble one another. Of course, this would mean also that one has an indirect duty to perform actions that reinforce duties to rational beings in this way. Showing respect for dead bodies would be one example.

---

[10] "Scandalum" is an example that furnishes an excuse to act badly.

But the analysis can be extended to the second role—where misconstrual enters into the picture. Once again, one should refrain from performing the action because it will lead to a lack of concern for rational beings. Of course, with the fur-coat case we get an "indirect" indirect duty. But in the case of the feminist we can see why he would be reluctant to use the toilet in a strip-tease joint. He might be mistaken as a patron, thus lending legitimacy to an establishment that unabashedly treats rational beings merely as means.

But Kant's analysis is not entirely satisfactory. First of all, it strikes me as arbitrary and misdirected. Indirect duties were postulated merely to explain our feelings or intuitions that happened to conflict with Kant's theory. Suppose, though, that we had a feeling that combing cat hair was terribly wrong. Would it then follow that we have an indirect duty to refrain from combing cat hair? Because, perhaps, it resembles pulling out people hair? I think, intuitively, that the reason we have misgivings about these actions is due to their consequences. And insofar as Kant's analysis is plausible it recognizes this. Thus, the analysis allows for back-door consequentialism. We are worried about how the perceptions of these actions will affect others.

Also, Kant's analysis may have difficulty handling other cases. Sometimes it is even wrong to do something that looks virtuous. Suppose you are concerned, as we all are, that children develop a healthy fear of strangers. Offering an unrelated child a ride home from school is not a good thing to do, even if your motives are pure. That is because the children witnessing the kindness may lose their fear of strangers.

I do not see how Kant's view could really deal with this case effectively. Would he argue that I have an indirect duty to avoid giving rides to school children because my action resembles the action of the potential child abuser? But how is anybody's sense of duty corrupted? My sense of duty is not corrupted, nor is the sense of duty any child witnessing my altruism. Rather, healthy instincts are not reinforced. The analysis here is purely consequentialist.

One might believe, once again in a Kantian vein, that what is morally dubious or questionable about these actions is that in performing them the agent deceives others and so manipulates them. Yet these actions are not actually deceptive because the agent does not intend to generate false beliefs. Indeed, if she could perform the action without generating those beliefs, she would. Still, in performing the action the agent does mislead others, and perhaps this is what explains the unease.

But performing misleading actions—and even deceptive ones—is not sufficient for moral blame. When Roger puts shoulder pads in his jacket he may mislead others about his build, but this is not immoral. The best answer is that what explains the unease is that people are worried about the consequences of performing mimetically immoral actions. The consequences of being misconstrued are various: (1) a bad example is set, and the agent stands in danger of corrupting himself and others; (2) the agent could give offense to others; or, as in the case of Pompeia, (3) the action could make the lives of others miserable. The split in intuitions occurs because the bad consequences are not the same as the ones that make the independently immoral act immoral.

That setting an example is a legitimate moral consideration is supported by the way we interact with children. When we interact with them we are especially careful in monitoring our conduct, because we recognize the importance of setting a good example (or at least refraining from setting a bad example). Children can very easily misunderstand the actions of others, because they are unaware of the relevant facts in a given situation. One must be careful when telling jokes around children, since often they cannot see the joke. One must be careful in telling a judicious lie around children, since they cannot tell the difference between the judicious lie and the mean lie. Anyone who did not take this into consideration when in the presence of children would be morally lax.

In most of our interactions with adults, relevant facts are clear, such as the assumptions and beliefs it would be reasonable for a person to make in a given clear context. Context usually is pretty clear for the purposes of moral evaluation. But not always. Thus, care needs to be taken because a bad example can and will have a bad effect on others. This consideration becomes stronger the more the agent is in the public eye; that is, the larger her audience is. A police officer, a judge, and a senator must avoid even the appearance of immorality.

But more ordinary examples abound as well. People will often feel constrained to make restitution for things that merely look immoral. A person walking in a store looking at vases may pay the management if a vase, coincidentally, should fall and break near her. Even though she did not break it, it looks as though she did, and if she failed to pay for it she could be setting a bad example. If the cost of the vase is small, she will think it worth paying to offset any negative effects.

Misconstrual can also lead to giving unnecessary offense to others. When I visit my grandmother, for example, I am careful not to take along any book that might have 'sex' in the title. Even though any such book in my possession would be a purely scholarly work, my grandmother probably would not realize this, and become distressed and offended.

I should also point out that some of these misconstruals will be quite reasonable. Take the fur-coat case. In this case it would not be unreasonable for the observer to form the belief that the coat is real fur. She would simply be reasoning through a straightforward statistical syllogism. 99% of all coats like that are fur, therefore that coat is fur.

But what if the misconstrual is unreasonable? What if it is due to a fallacy to which humans are particularly prone? For example, people are very prone to committing the gambler's fallacy. Suppose I am a captain trying to regroup my men in the midst of a battle. I jump into a nearby crater to try to protect myself from shrapnel, and I urge my fellow soldiers to seek shelter as well. Well, another soldier might think me a greedy selfish pig for availing myself of the crater, when in fact I am not. That is because a lot of people think that the chance of a bomb falling on that crater is less than it falling on some other spot. After all, is not the probability of two shells falling on the same spot very low? So, my action looks selfish, and this might serve to demoralize my soldiers. But it only looks selfish because people are prone to commit the gambler's fallacy. Lying in that crater to protect myself from shrapnel, I have the same chance as anybody else of getting hit by a bomb.

People are also remarkable for their superstitious tendencies. Suppose, as I prepare to run in a marathon, the computer arranging the shirt numbers assigns 666 to me. If I wear that number I may look like a devil worshiper, when in fact I am not. And that shirt may strike fear into the superstitious hearts of many people. .

Some might argue that the agent is only responsible for reasonable misconstruals that others make—not the unreasonable ones. A consequentialist will argue, however, that this distinction is basically irrelevant since, as long as the misunderstanding can be predicted, it must figure into the agent's calculations. The agent is, of course, not responsible for the odd irregularities of others, since she has no way of predicting them.

I must say at this point that I am not arguing that people must at all times display behavior conforming to accepted moral norms. I should hope that the account could allow for the moral rebel here and there. I am not saying categorically that it is wrong to wear very good fake furs, because long-term consequences must be taken into consideration. If wearing fur is immoral, but some people are deeply attached to fur, so attached that without a good substitute they would keep on wearing the real thing, then fake fur could be a happy compromise. Maybe at first bad values are reinforced because of the resemblance. But as people come to accept the fake fur on a more widespread basis, this effect will disappear, because people will recognize the fake as a fake. Thus, overall, the consequences of wearing these coats would be good. This is fine, as long as one did expect the long-term consequences to work out. But this itself points to the force of resemblance—it still has to be taken into consideration by the moral agent. During the transition period, before the effect of the resemblance is neutralized by familiarity, people should take steps to make sure that their actions are not misinterpreted. Maybe, in this case, by wearing "Great Fake" buttons on one's coat. Another example: Jewish law forbids the consumption of dairy products with meat. Thus, when nondairy creamer was developed, one would have thought that this was a boon to the dinner diet at Jewish tables. The nondairy creamer looked just like cream, however, so people were too embarassed to use it for a long time. Some were worried that others would quite naturally mistake it for the real thing. What some people did to get around this was to serve the nondairy creamer in its carton—even though this was tacky looking.

IV CONCLUSION

My account of mimetically immoral (and moral) actions is obviously consequentialist. But that is because it is clear that what makes mimetically immoral acts immoral is their consequences. There is nothing intrinsically wrong, or immoral, with these actions. This analysis bolsters consequentialist ethics because that approach *leads* one to the correct analysis of these actions. The spirit of this approach runs counter to deontology which famously maintains that immorality is not exhausted by consequences. Of course, deontologists do not disregard consequences altogether—but they maintain that by and large other factors matter morally as well. Thus, a deontologist could agree with my analysis of mimetic immoralities by maintaining that consequences are for these actions what determines their moral quality. Still, by being the fruitful theory here, consequentialism gains some ground. Rather than lead us to a plausible analysis, deontology merely responds to cases as they crop up.

To return to Caesar's wife, Pompeia: if she believed that her actions could be misconstrued in such a way as to cause harm to her husband, or to corrupt others by her example, she had a moral reason not to do them. Being good is sometimes a matter of looking good.

JULIA DRIVER

Brooklyn College/CUNY

SEMANTIC SUPERVENIENCE AND REFERENTIAL INDETERMINACY*

MY topic in this paper is the bearing of semantic supervenience (the thesis that semantic facts supervene on nonsemantic facts) on currently influential arguments for the indeterminacy of reference. I shall argue that semantic supervenience is a two-edged sword: it is an essential weapon in some attacks on referential determinacy (e.g., that of Saul Kripke's Wittgenstein), but it may also be used to undercut the main points against referential determinacy which have been urged by Hilary Putnam. This is because semantic supervenience leads to semantic externalism, a doctrine I shall expound and defend by analogy with externalism in epistemology.

I. ARGUMENTS FOR INDETERMINACY

It will be instructive to begin by considering the general form of many arguments for indeterminacy.

1. Supervenience: an *A*-statement, if true or false, is true or false in virtue of some *B*-fact.
2 Underdetermination: the totality of *B*-facts does not suffice to determine a truth value for certain *A*-statements.

Therefore,

3. Indeterminacy: some *A*-statements lack determinate truth value

The two premises may seem at first glance to be inconsistent with each other. Supervenience (or rather, its inverse) sometimes goes by the name 'determination'; to say that one stratum of truth or fact supervenes on another is to say that the latter *determines* the former. So how can supervenience be combined with *under*determination? The answer is that it can be so long as one does not think of supervenience as automatically involving bivalence for the supervenient domain. Standard definitions of supervenience leave the question of bivalence open; they simply require that truths in the supervenient domain, to whatever extent there *are* truths in that domain, have determining grounds or sufficient conditions in the

---

* I am indebted to Felicia Ackerman, Dorit Bar-On, Simon Blackburn, John Heil, Eli Hirsch, Jaegwon Kim, Michael Resnik, Ernest Sosa, Carl Posy, Peter van Inwagen, a colloquium audience at the University of North Carolina/Chapel Hill, and the National Humanities Center.

0022-362X/92/8907/344–61

subvenient domain.[1] When we combine supervenience with underdetermination, the result is a failure of bivalence, which is precisely what is involved in most varieties of indeterminacy.[2]

Here are some familiar arguments that exemplify the pattern above:

Indeterminacy of identity. Letting the *A*-statements be statements of personal identity and the *B*-facts be facts about physical and psychological continuity, we can see Derek Parfit[3] as arguing from 1 and 2 as premises to 3 as conclusion. In some of the puzzle cases he discusses, he maintains that the question 'Is a certain person who emerges at the end of process *P* identical with the person who entered it?' has no correct answer. Other parties to the debate can be seen as denying one of the premises. Roderick Chisholm and Kripke deny 1—they believe that identity is a "Further Fact"; Bernard Williams and Sydney Shoemaker deny 2—they espouse reductionism without indeterminacy.

Indeterminacy of the past. Now let the *A*-statements be statements in the past tense and the *B*-facts be facts about present memories (or whatever else is admitted as currently available evidence bearing on the past). Michael Dummett[4] has flirted with a position that again argues from 1 and 2 to 3: there is no truth about the past without present evidence to warrant its assertion; there is no present evidence to warrant asserting that Caesar did, nor any to warrant asserting that he did not, eat eggs on the day he crossed the Rubicon; therefore, there is no truth either of his having eaten eggs that day or of his not having done so.

---

[1] To forestall confusion, I could perhaps use a label such as 'conditional supervenience' for premise 1. But that would be a redundancy; supervenience as standardly understood is already conditional. See Kim, "Concepts of Supervenience," *Philosophy and Phenomenological Research*, XLV, 2 (December 1984). 153–76, esp. the definitions on pp. 163, 165.

[2] One might wonder whether the validity of the argument from supervenience and underdetermination to indeterminacy does not depend on a further premise ruling out "multiple base" supervenience, i.e., the supervenience of the *A* domain on two distinct domains, *B* and *C*. The idea would be that, even if *A* is underdetermined by *B*, it might be fully determined by *C*, and thus not indeterminate. But this idea turns out to be incompatible with my two premises as already stated If *A* is underdetermined by *B*, some *A*-statements are not rendered true or false by any *B*-facts; so if these statements are determinately true or false nonetheless, they cannot be true or false in virtue of *B*-facts; but that contradicts the premise that *A* supervenes on *B*. So my two premises are enough by themselves to yield indeterminacy. There is no need to rule out multiple base supervenience, since the premises already imply that, if *A* supervenes on any further domain *C*, it must be determined by *C* no further than it is determined by *B*. (I thank Kim for getting me to think about this point.)

[3] *Reasons and Persons* (New York: Oxford, 1986), pt. 3.

[4] "The Reality of the Past," in *Truth and Other Enigmas* (Cambridge: Harvard, 1978), pp. 358–74. I simplify somewhat, but I think this argument captures the gist of one position considered by Dummett.

Indeterminacy of position/momentum. Quantum mechanics does not allow us simultaneously to assign exact position and exact momentum to a particle; one or the other of these two values must, in any given case, be underdetermined relative to the total QM state description. If we now add (what Einstein was never willing to admit) that QM is "complete," in the sense that it tells us the whole story about what is going on at the microlevel, the result is indeterminacy: a particle that has determinate momentum will lack determinate position, and a particle that has determinate position will lack determinate momentum. Once more, we have argued our way to indeterminacy from a combination of premises that amounts to supervenience plus underdetermination.

Indeterminacy of reference. To come at last to the topic of this paper, I believe that W.V. Quine, Kripke's Wittgenstein, and others can be viewed as using an argument of the form I have displayed to establish the indeterminacy of reference. The *A*-statements in this case are statements of the form 'term *x* refers to object *y*'. The *B*-facts, for Quine, are facts about verbal behavior and dispositions to verbal behavior; for Kripke's Wittgenstein, they are naturalistic facts generally, including all the mental facts you like. The two philosophers I have mentioned may actually use a strengthened form of the argument: in *no* case do the *B*-facts determine a truth value for an *A*-statement, so there is *never* a fact of a term's referring to one thing rather than another.

Quine's argument for the indeterminacy of translation may also be viewed as having the form above. Quine's formulations of the indeterminacy of translation often misleadingly equate it with the *underdetermination* of correct translation by the totality of verbal behavior and behavioral dispositions.[5] But such underdetermination cannot be all there is to Quinean indeterminacy. If it were, Quine would have no warrant for saying, as he often does, that there is no fact of the matter regarding which of two divergent translation manuals is correct; it would be open to others to maintain that there is indeed a fact of the matter, holding, perhaps, in virtue of mental states not manifested in behavior, or perhaps even holding simply as an ulterior "further fact." To rule these positions out of court, Quine must insist that linguistic meaning supervenes on verbal behavior. This is exactly what he does insist upon: "There is nothing in linguistic meaning, then, beyond what is to be gleaned from overt behavior in observable circumstances."[6] Quine also maintains that scientific truth in general is underdetermined by the totality of observable facts, but in this case he does *not* espouse the corresponding supervenience thesis; that is why he is free to hold that there is no indeterminacy in our overall theory of nature.[7] An important moral here, generalizable to other problem areas in philosophy as well, is that underdetermination alone does not get you indeterminacy; for that you need underdetermination plus supervenience.

---

[5] See, e.g., *Word and Object* (Cambridge: MIT, 1960), pp. 27, 73.

II. PUTNAM'S MODEL-THEORETIC ARGUMENT AGAINST REALISM

The indeterminacy of reference plays a crucial role in Putnam's so-called model-theoretic argument against metaphysical realism. The tenet of metaphysical realism under attack in this argument is that truth is "radically nonepistemic," by which Putnam means that even an "epistemically ideal" theory could be false. The essentials of the argument may for present purposes be set forth as follows:

1. Suppose a theory *T* is epistemically ideal, i.e., meets all operational and theoretical constraints, but is false.
2. Then there is an interpretation *I* under which *T* is true.
3. The metaphysical realist is in no position to say that *I* is not an intended interpretation of *T*.
4. So the metaphysical realist cannot say after all that *T*, an ideal theory, is false.[8]

Let me note three points briefly before getting on to the main topic. (1) I have omitted mention of the model-theoretic considerations that give the argument its name. Putnam invokes them in passing from 1 to 2, but they do not loom quite so large as his exposition would suggest; nor do they raise what I see as the central issues in the argument.[9] (2) The conclusion as formulated above is the weakest conclusion supported by the premises which is incompatible with metaphysical realism as Putnam defines it.[10] (3) To what extent is this argument an argument against *realism?* The truth of sentences is a function of two things, their interpretation and how the world is. What the foregoing argument shows, if successful, is that interpretation is indeterminate (at least for the metaphysical realist) and that sentential truth is therefore indeterminate (at least for the metaphysical realist). Even if Putnam is right about this, it would not follow in the least that *how the world is* is indeterminate. A core element of realism therefore remains untouched by Putnam's argument.[11]

---

[6] Quine, "Indeterminacy of Translation Again," this JOURNAL, LXXXIV, 1 (January 1987): 5–10, at p. 5.

[7] These points of Quine interpretation are confirmed by "Indeterminacy of Translation Again."

[8] Sources of the argument in Putnam are "Realism and Reason," *Proceedings of the American Philosophical Association*, L (1977): 483–98, repr. in his *Meaning and the Moral Sciences* (New York: Routledge & Kegan Paul, 1978), pp. 123–40, hereafter MMS, *Reason, Truth, and History* (New York: Cambridge, 1981), ch. 2, hereafter RTH; and "Models and Reality," in Putnam, *Realism and Reason* (New York: Cambridge, 1983), pp. 1–25, hereafter RR. Interpretations of the argument similar to what follows are offered in Alvin Plantinga, "How to Be an Anti-Realist," *Proceedings and Addresses of the American Philosophical Association*, LVI (1982): 47–70; Anthony L. Brueckner, "Putnam's Model-Theoretic Argument against Metaphysical Realism," *Analysis*, XLIV, 3 (June 1984): 134–40; and Mark Heller, "Putnam, Reference, and Realism," in *Midwest Studies in Philosophy*, XII, Peter A. French, Theodore E Uehling, Jr, and Howard K. Wettstein, eds. (Minneapolis: Minnesota UP, 1988), pp. 113–27.

I turn now to the main issue. The heart of Putnam's argument is clearly premise 3, the premise in which referential indeterminacy comes to the fore. Why is anyone compelled to accept this premise? What prevents the metaphysical realist from saying that I (the concocted interpretation under which T is true) is an incorrect interpretation? In support of the premise, Putnam claims that I satisfies all operational and theoretical constraints. But many philosophers will invoke constraints on reference that go beyond those admitted by Putnam. In particular, many will invoke *causal* constraints—a term t as used by S is correctly interpretable as referring to Ks only if there are appropriate causal connections between Ks and S's use of the term. More generally, some will contend that reference is to be analyzed somehow in terms of a naturalistic relation R, which may include causal relations and other ingredients as well:

x refers to y if and only if x bears R to y.

This is the schema for a proposal of Hartry Field's as Putnam introduces it (RTH, 45).

Putnam is well aware of this possibility and says the following things by way of objection to it:

---

[9] As David Lewis points out, without using the Loewenheim-Skolem theorem, one can already show that any large enough world can satisfy any consistent theory What the Loewenheim-Skolem theorem adds is simply that any denumerable world is large enough. See Lewis, "Putnam's Paradox," *Australasian Journal of Philosophy*, LXII, 3 (September 1984). 221–36, at p. 229.

[10] In a longer version of this paper, I take up the question whether the argument can be made to yield the stronger conclusion that an ideal theory *must be true*.

[11] van Inwagen makes a similar point about Putnam's argument concerning brains in vats, see his "On Always Being Wrong," *Midwest Studies in Philosophy*, XII, pp. 95–111. According to van Inwagen, Putnam's arguments have everything to do with what it takes for a person to be intentionally related to a proposition and nothing to do with what it takes for a proposition to be true. Compare also Lewis, p. 232 "The *metaphysics* of realism survives unscathed."

> A causal theory of reference would not help, "for how 'causes' can uniquely refer is as much of a puzzle as how 'cat' can, on the metaphysical realist picture" (MMS, 126); "the reference of 'x bears R to y' is itself indeterminate" (RTH, 45); "it is a puzzle how we could *learn to express* what Field wants us to say" (RTH, 46)

This is the "just more theory" objection. If the interpretation of T was indeterminate before, it will remain indeterminate if we add Field's account of reference to T, for the account itself suffers from the same indeterminacy. The theory-cum-account has interpretations unintended by Field that make it true, but there is no way to rule these interpretations out of court—there is nothing to make it the case that they are "really" unintended.

Several of Putnam's critics have thought that the foregoing objection to the causal theory is completely wrongheaded. They have pointed out that in putting forth the causal theory, one does not "just add more theory"; one poses an *external* constraint on reference, a constraint that works from outside the theory rather than within it. Thus David Lewis writes:

> The constraint is *not* that an intended interpretation must somehow make our account of C [the constraint] come true. The constraint is that an intended interpretation must conform to C itself (*op. cit.*, p 225).

It is not causal language that fixes reference; it is causation itself.

To the exasperation of his critics, Putnam has refused to see this point. He reports it accurately enough, but then more or less simply repeats what he has said before.[12] I hold no brief for the causal theory, but I believe the critics are essentially right. Moreover, it is important to see that they are right, since the ploy used by Putnam against the causal theory could be used against *any* external constraint on reference, thereby dooming us to accept Putnam's own conclusions—conclusions that many find incredible. In the hope of getting more people to see the externalists' point, perhaps including Putnam himself, I propose to pursue some analogies between epistemology and semantics.

III EPISTEMOLOGY AND SEMANTICS

One of the major tasks of epistemology is to articulate *epistemic principles*—principles of the form 'If ______, then subject S is justified in believing proposition p'. Such principles divide into two classes. One class includes principles that warrant inference from already justified propositions to further propositions; the antecedents of such principles will specify that certain propositions already have some epistemic status for the subject. But not all epistemic principles can be like this. There must also be a class of epistemic principles that specify the nonepistemic conditions under which some beliefs come to have some epistemic status or other in the first place—the conditions, one might say, under which epistemic status is *generated*. Examples of principles belonging to this second class, which I shall call *generation principles*, are Descartes's principle that whatever is perceived clearly and distinctly is certain, Chisholm's principle that whatever is self-presenting is evident, Alvin Goldman's principle that a belief formed as the result of a reliable process is justified, and the coherentist's principle that a belief belonging to a suitably coherent set of beliefs is justified. As the last two examples should make clear, there is nothing inherently foundationalist in the requirement that a theory of knowledge include at least one generation principle. The requirement is really just the requirement of epistemic supervenience—that there be some nonepistemic features that ultimately underlie the instantiation of any epistemic property.[13]

---

[12] See RR, introduction, esp. xi, and 295–6.

A corollary of epistemic supervenience is *epistemological externalism:* the view that there are factors that contribute to knowledge and make it possible regardless of whether they themselves are known. Causal and reliability theories of knowledge are perhaps the best known externalist theories in epistemology, but it is worth pointing out that even a theory like Descartes's is, in an important sense, externalist. For Descartes's theory specifies a condition that gives rise to knowledge—namely, clear and distinct perception—regardless of whether it itself is known to obtain. One way to see the externalist implications of epistemic supervenience is as follows: a generation principle says that the obtaining of whatever condition is specified in its antecedent is *sufficient* for the accruing of whatever epistemic status is specified in its consequent (knowledge, justification, or what have you). That means that nothing *else* is necessary; in particular, knowledge that the condition obtains is not necessary.[14] Thus, a Cartesian knower does not have to know that his state is one of clear and distinct perception; Goldman's knowers do not have to know about the reliability of their own cognitive processes, and so on.

---

[13] For further discussion of generation principles, see my "Foundationalism, Epistemic Principles, and the Cartesian Circle," *The Philosophical Review*, LXXXVIII, 1 (January 1979): 55–91 For more on epistemic supervenience, see Sosa, "The Foundations of Foundationalism," *Noûs*, XIV, 4 (November 1980): 547–65, as well as various of the earlier and later essays in *Knowledge in Perspective* (New York: Cambridge, 1991); Alvin Goldman, "What is Epistemic Justification?," in *Justification and Knowledge*, George S. Pappas, ed. (Boston: Reidel, 1979), pp 1–23; and my "Epistemic Supervenience and the Circle of Belief," *The Monist*, LXVIII, 1 (January 1985): 90–104.

The reader is likely by now to have guessed where I am heading. Semantic principles in the theory of meaning or reference are in important ways analogous to epistemic principles in epistemology. Field's schema is precisely the schema for a generation principle in semantics: its right-hand side should be seen as spelling out in nonsemantic terms the conditions under which certain semantic relations first arise.[15] If semantic phenomena, like epistemic phenomena, are supervenient on natural phenomena, then there must be at least one true principle satisfying Field's schema—a causal principle, as it might be, but if not that, some other principle specifying the natural basis for reference.

Let us suppose, then, that the following is a true generation principle in semantics (I now ignore the other half of Field's biconditional): if x bears R to y, then x refers to y. In light of our discussion above, it should be clear that, if this principle is true, all that has to happen in order for x to refer to y is that x bear R to y. Nothing more is necessary; in particular, it is not necessary that we be able to *say* "what Field wants us to say." Contrary to what Putnam repeatedly insists upon,[16] we do not have to be able to "single out" one relation as the intended referent of 'R' before the causal theory (or the R-theory) can do its job. We can do the singling out later, as I shall show in the next paragraph, but we need not do it as a precondition of the successful working of the theory. One might put the point this way: as knowledge makers need not be known, so reference makers need not be referred to.[17]

---

[14] More accurately: nothing is necessary for the consequent beyond what is necessary for the antecedent itself. It is quite plausibly arguable in the case of many generation principles that their antecedents may obtain without the subject's having to know that they do.

[15] That Field himself sees the schema in this light is clear from Field, "Conventionalism, and Instrumentalism in Semantics," *Noûs*, IX, 4 (November 1975): 375–405, at p 382.

[16] See, e g., "Why There Isn't a Ready-Made World," in RR, 205–28; and RTH, 46

[17] For further discussion of the bearing of externalist theories against Putnam's views, see Heil, "The Epistemic Route to Anti-Realism," *Australasian Journal of Philosophy*, LXVI, 2 (June 1988): 161–73; and Heller, *op. cit.* Heil and Heller emphasize epistemological externalism rather than semantic externalism.

The analogy with epistemology can be carried one step further. Knowledge makers *need* not be known, I just said, but they *can* be. This will happen whenever knowledge-making states of affairs come themselves to instantiate the antecedents of epistemic principles. If we should acquire clear and distinct perception about the occurrence of clear and distinct perception, or reliably formed belief about the reliability of the way in which a given belief was formed, we would come thereby to have knowledge of the occurrence of a knowledge maker. I have argued elsewhere that there is nothing objectionably circular about this.[18] Similarly (as several writers have pointed out), if our use of the term '*R*' were related by *R* itself to instances of *R*, that would make it the case that '*R*' refers to *R*.[19] There is nothing objectionably circular about this, either; as Michael Devitt points out (*ibid.*), there is no reason why a comprehensive theory of reference cannot (and should not) apply to its own words.

Putnam is aware of the possibility of holding that '*R*' refers to *R* in virtue of self-subsumption (as I shall call the circumstance in which '*R*' stands in *R* to instances of *R*). Leaving aside an objection based on the theory of types, he objects to self-subsumption chiefly on the ground that *many* putative reference relations are similarly self-subsuming. We can cook up a crazy scheme of reference under which 'tables' refers to alligators. To go with it, we can concoct a gerrymandered relation causation* such that our use of the word 'tables', besides being causally related to tables, is also causally* related to alligators. (See RR, introduction and 295–6). Moreover, things may even be so rigged that our use of 'cause*' (if we do use it!) is related by causation* to instances of itself. So on what ground could we say that causation and not causation* is what makes for reference, and hence that 'tables' refers to tables rather than to alligators? Evidently, self-subsumption is not enough.

At this point, it will again be instructive to consider a parallel situation in epistemology. I have argued elsewhere that there need not be any vicious circularity in arriving at knowledge of epistemic principles by a route that uses those very principles. For example, the layman's reason for believing in induction—"it works"—can be established inductively and serves to justify induction itself.[20] A natural objection to this proposal is that it could be used with equal right by counterinductivists to justify counterinduction. Counterinductivists believe that, if $x\%$ of the *A*s have been *B*s in the past, one should believe that $(100 - x)\%$ of the *A*s will be *B*s in the future. Counterinduction has had a miserable record of success in the past, but a counterinductivist would simply regard this as evidence that it will work splendidly in the future. He can cite his own principle in self-defense just as surely as the inductivist, but it is absurd (so the objection goes) to think that this amounts to any sort of justification of it. I trust that the parallel with Putnam's point about causation* will be clear.

---

[18] "Foundationalism, Epistemic Principles, and the Cartesian Circle."

[19] See, e.g., Brueckner, p. 138, and Michael Devitt, *Realism and Truth* (Princeton: University Press, 1984), pp. 189–90.

[20] See my "Reliability, Justification, and the Problem of Induction," *Midwest Studies in Philosophy*, IX, Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds (Minneapolis: Minnesota UP, 1984), pp. 555–67.

I believe there is a good reply to this objection, whether taken in its epistemological or its semantical version. Supposing our world to be one in which induction is reliable (and counterinduction not), and supposing that reliability rather than its opposite is what confers justification, induction will be justified and counterinduction will not be, even though both are self-subsuming. It is not ***mere*** self-subsumability that confers justification, but self-subsumability under a true principle. (Which is not to say a principle ***known*** to be true; the principle can state an external condition of justification.) If it is true that reliable beliefs are justified beliefs and that induction is reliable, then counterinductivists will be out of luck. Similarly, if it is a fact that causation has something important to do with reference but causation* does not, then causal* theorists will be out of luck. Their theory will be mistaken despite being self-subsuming.

In the last paragraph, I have drawn out an important consequence of the putative fact that causation but not causation* makes for reference. Putnam seems at times to doubt that there could be any such fact. In an article replying to Field, he says that to hold that reference is fixed by causation rather than by causation* "because that's the ***nature*** of reference" is ***unabashed essentialism***.[21] And elsewhere in response to Field, he claims that the fact, if it is a fact, that *R* underlies reference must be taken as "a kind of primitive, surd, metaphysical truth" (RTH, 46).

I believe that Putnam's claim against Field should be compared with the following: the fact that it is reliability (and not, say, a success rate of twenty, ten, or zero per cent) that makes for justification must be taken as a primitive, surd, metaphysical truth. Primitive and metaphysical it may be; whether it is also surd I leave to the reader's own consideration.[22]

---

[21] Putnam, "Reply to Two Realists," this JOURNAL, LXXIX, 10 (October 1982): 575–7.

[22] A similar point is made by Brueckner, p. 138.

IV ANOTHER ARGUMENT FOR REFERENTIAL INDETERMINACY

There is one more possible point in Putnam's defense I would like to consider. Ernest LePore and Barry Loewer[23] defend Putnam against some of Devitt's criticisms by citing a Putnam doctrine that I have so far said nothing about:

> Devitt has missed the force of Putnam's argument that there are alternative reference relations that result in identical assignments of truth values at every possible world. If there is a causal characterization of reference, there also are alternative characterizations (in terms of cause*, cause**, and so on) that agree on the truth values of statements at all possible worlds. The question Putnam asks the MR is why say that the account of reference in terms of cause rather than in terms of cause* is the correct one (*ibid.*, p. 464)?

The Putnam-endorsed point to which LePore and Loewer are calling attention is this: the correct scheme of reference for a language is underdetermined by the distribution of truth values it yields for all sentences of the language in all possible worlds (since competing schemes can yield identical distributions). Having noted this, they go on to ask the supposedly rhetorical question, what could make one rather than another of the competing schemes correct? If their own answer (or Putnam's) is 'nothing', note what they (or he) must be presupposing: that the reference of terms supervenes on the truth values of sentences containing them. (Recall a lesson from section I: underdetermination by itself is insufficient to beget indeterminacy; to get from underdetermination to indeterminacy, you must assume a supervenience thesis as well.) The needed supervenience thesis could perhaps be formulated as follows: two schemes of reference for a language *L* do not differ in correctness unless they also differ in the truth values they yield for at least one sentence of *L* in at least one possible world.[24] If this thesis is granted, the indeterminacy of reference swiftly follows.

But why on earth should anyone accept the thesis? I would like to make three points about it. (i) It is obviously not a thesis that causal theorists are going to accept. If they do not accept it, they can cheerfully admit Putnam's underdetermination point. (ii) To claim (as the thesis does) that reference is exhausted by its contributions to truth conditions is surely to indulge in what Putnam once disparaged as "negative essentialism": it is to claim that certain constraints exhaust the content of a given concept even though they do not suffice to determine its extension.[25] Why is negative essentialism about reference supposed to be less objectionable than the "unabashed essentialism" Putnam complains of when causal theorists say it is the nature of reference to be founded on causation? (iii) Finally, how plausible is it anyway to suppose that reference supervenes on nothing but the distribution of truth values (in all possible worlds) over the sentences of a language? To my mind, that supposition is every bit as arbitrary as the following one: reference supervenes on nothing but the *ratio* of true sentences to false ones. If two schemes of reference for a language *L* both make half of its sentences true and half false (and there are few that do not do this!), then they are equally correct.

---

[23] "A Putnam's Progress," in *Midwest Studies in Philosophy*, XII, pp 459–73.

[24] This is very similar to the principle John Wallace expounds and attributes to Frege in "Only in the Context of a Sentence Do Words Have Any Meaning," in *Midwest Studies in Philosophy*, II, Peter A French, Theodore E. Uehling, Jr., and Howard K Wettstein, eds. (Minneapolis: Minnesota UP, 1977), pp 144–64. The principle I formulate would be equivalent to the Wallace-Frege principle if the "truth conditions" in Wallace's formulation were given by assignments of truth values in all possible worlds.

V ALTERNATIVES TO SEMANTIC SUPERVENIENCE

We have seen so far that, if the thesis of semantic supervenience is correct, Putnam's objections to causal theories of reference are indeed ill-conceived, just as his critics contend. Well, then, why could he not simply deny such supervenience? In this section, I would like to explore the options available to anyone who does deny it. I believe that once we are clear about what the alternatives are, the doctrine of semantic supervenience will appear in a better light.

To affirm semantic supervenience is to affirm the following: necessarily, if any semantic fact ever obtains (e.g., that x refers to y, or that sentence S1 means the same as S2), it does so in virtue of some nonsemantic fact. To deny semantic supervenience would therefore be to embrace one of the following alternatives as possible: either (i) some semantic facts are not made true by *any* further facts, or (ii) they hold in virtue of further facts, but always facts that are themselves semantic. Let us consider each alternative.

Alternative (i) is accepted by some philosophers, but would certainly not be acceptable to Putnam. Brute semantic facts, or semantic facts that do not hold in virtue of any deeper stratum of fact, are precisely his target when he inveighs against "magical" theories of reference.[26]

---

[25] Putnam, *Mind, Language and Reality* (New York: Cambridge, 1975), pp. 161–4

[26] This is clear from the dialectical contexts in which Putnam levels the charge of magic in the first two chapters of RTH. The *terms* he uses to explain what he

Alternative (ii) calls to mind C. S. Peirce's strange doctrine of signs and interpretants. According to Peirce, a *sign* signifies an *object* only through the mediation of a third thing, an *interpretant;* and *every interpretant is itself a sign,* signifying an object. These premises obviously generate an infinite regress (or perhaps a circle, if it is allowed that an interpretant may have for its own interpretant a sign that has already occurred earlier in the series). There is an echo of Peirce's view in Putnam's remark that "any concept we may have of the *relation* between a sign and its object is *another sign*" (MMS, 126–7).[27] Nonetheless, I cannot really believe that Putnam would wish to accept an infinite regress of interpretants. Peirce's regress is surely a *reductio ad absurdum* of the premises that generate it.

Here, then, is why Putnam should accept semantic supervenience: if he denies it, he must believe either in the very magic he deprecates or in something worse, namely, that reference is begotten only by reference, either *ad infinitum* or in a circle.[28]

The alternatives before us can be expanded and illustrated further by more analogies from epistemology. Consider the familiar problem of the infinite justificatory regress, which arises if we accept the following four assumptions:

1. Some beliefs are justified.
2. If a belief is justified, there is some further fact in virtue of which it is justified.
3. That in virtue of which a belief is justified must itself either be or involve a justified belief.
4. There cannot be circles in justification. (I shall not attempt to formulate this assumption precisely, but the need for it is clear.)

These assumptions are readily seen to generate an infinite regress of justification. Confronted with such a regress, we have five options. (i) We could accept all four assumptions and the implied regress. Peirce is sometimes said to have embraced the possibility of

---

means by magic—it is the view that the connection between signs and objects (or between representations and what they represent) is *intrinsic* (RTH, 5, 47, 51) or *necessary* (RTH, 3, 17–18, 51)—are not aptly chosen. A primitive relation (his real target) need not be intrinsic, nor vice versa, nor need either of these be necessary

[27] Putnam connects the point with Occam rather than Peirce, the Peircean overtones are somewhat muffled by his use of the word 'concept'.

[28] Strictly speaking, he need only believe in the *possibility* of these things, but what would be absurd as actuality is equally absurd as possibility.

such a regress,[29] but few other philosophers have taken it seriously. (ii) We could deny assumption 1, landing ourselves in total, unmitigated skepticism. (iii) We could deny assumption 2, positing self-justifying or intrinsically justified beliefs. (iv) We could deny assumption 3. As I shall explain more fully in a moment, that would mean embracing epistemological externalism, at least of the minimal variety implicated in epistemic supervenience. (v) Finally, we could deny assumption 4. That is the option sometimes associated with coherentist epistemologies.

In my view, we must in any case deny assumption 3, thus embracing epistemic supervenience. In defense of this stand, I would point out that positions that apparently deny one of the other assumptions are better construed as denying 3 instead (or perhaps in addition). For example, the most plausible coherentism is not the linear coherentism that says belief *A* is justified because belief *B* is justified, which is justified in its turn because belief *C* is justified, and so on, until we get back around to *A;* it is the global coherentism that says each belief is justified in virtue of its place in the entire system.[30] (Light dawns gradually, or perhaps in a flash, over the whole.) On this picture, justification supervenes upon coherence relations that are thoroughly nonepistemic, neither involving knowledge nor themselves needing to be known. Again, the most plausible foundationalism does not posit beliefs that are literally self-justifying; it posits beliefs that are justified by further facts—an option that does not compromise foundationalism so long as the facts in question are nondoxastic and nonepistemic.[31] It thus denies assumption 3.

Let us now turn to semantics, which I believe may be viewed in quite parallel fashion. As before, there are four assumptions that generate an infinite regress and five resulting options.

1. There is such a phenomenon as reference—there are truths of the form 'term *x* refers to object *y*' (for subject *S* or in language *L*, of course).

---

[29] See "Questions Concerning Certain Faculties Claimed for Man," in *Writings of Charles S. Peirce*, Vol. 2, Edward C. Moore, ed (Bloomington Indiana UP, 1984), pp. 193–211. What he actually defends is the possibility that each cognition of an object be "determined" by an earlier cognition; it is not so clear that he would endorse an infinite regress of *justification*.

[30] I take the terms 'linear' and 'global' from Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge: Harvard, 1985), pp. 89–93, though not quite in his senses. My global coherentist accepts epistemic supervenience; BonJour's evidently does not For more on the most plausible coherentism as involving supervenience, see Sosa, "The Raft and the Pyramid," in *Knowledge in Perspective*, pp. 165–91, esp. sects. 7 and 8.

[31] I say more about this option in "Epistemic Supervenience and the Circle of Belief."

2. If one thing refers to another, there is some further fact in virtue of which it so refers.
3. That in virtue of which a sign refers must itself be a fact that either involves reference or is itself referred to.
4. There cannot be circles in reference (it cannot be that x refers as it does only because something else refers as it does and conversely).

What are the options this time?

(i) We could accept all four assumptions and the resulting infinite regress. This is the view to which Peirce was apparently committed, but he himself said the regress was absurd,[32] as I trust most readers will agree. I shall mention an alternative interpretation of Peirce presently.

(ii) We could deny assumption 1 and be thoroughgoing nihilists about reference. I think this is implicitly the position of Putnam and Quine, even though they prefer to portray their views less starkly, Quine allowing for "relative reference" and Putnam for "internal reference."[33]

(iii) We could deny assumption 2 and hold that the reference of some items to others is an ultimate, irreducible fact. If the referring items in question are words, we would have the position of Plato's Cratylus. If, more plausibly, it is held that words refer only through the mediation of thoughts, but that the reference of thoughts is primitive and irreducible, we would have the position of Chisholm and Franz Brentano. This, of course, is the position Putnam castigates as magic.[34]

(iv) We could deny assumption 3 and embrace semantic supervenience. This, as we have seen, is where the causal theory of reference fits into the picture.

(v) Finally, we could deny assumption 4 and embrace some sort of coherence theory of reference. As in the epistemological case, however, the most plausible coherence theories will deny not 4, but 3. To deny 4—to hold that each of two terms has the referential properties it does only because of the referential properties possessed by the other—really would be absurd. A more plausible coherence theory of reference would deny 3, holding that reference accrues to a system of terms when they are related to each other in ways describable without using semantic vocabulary. This may have been the view that Peirce actually held; it is at any rate the sort of view that A. J. Ayer attributes to him.[35] It is also arguable, though I shall not argue it here, that Kant's view about the reference of mental representations to objects fits here.[36]

---

[32] My source on Peirce is A. J. Ayer, *The Origins of Pragmatism* (San Francisco: Freeman & Cooper, 1968), p. 123.

[33] I say more about the surrogates for reference offered by Putnam and Quine in "The Myth of Relative Reference," unpublished

[34] On p. 110 of *Representation and Reality* (Cambridge: MIT, 1988), Putnam says that Chisholm and Brentano are right to regard intentional properties as primitive. So does Putnam now believe in magic? More likely, he only agrees with Chisholm and Brentano on the negative point that intentional properties are irreducible—not also on the positive point that they are features of the universe in good objective standing.

Let us now take stock. We have canvassed five options about reference, broadly construed to cover both linguistic reference and intentionality: circle, regress, magic, supervenience, and elimination. The circle and the regress can be lumped together as both involving the idea that reference is born only of reference—not a winning proposition. My own preference is for semantic supervenience, but if that proved an illusory hope, I would sooner believe in magic than embrace any of the other alternatives. What about Putnam? Magic he rejects, as we well know. I believe he would also reject the circle and the regress; see, for example, RTH, 43, where he objects to explaining reference in terms of intention on grounds of circularity. He does not take an explicit stand on semantic supervenience, but his dismissal of external constraints on reference depends, as I have argued, on denying or ignoring the implications of that doctrine. That leaves us with nihilism, or the elimination of reference, which I think is the ultimate upshot of Putnam's views.[37]

VI. CONCLUDING COMPARISON: KRIPKE'S WITTGENSTEIN

By way of conclusion, I would like to compare Putnam's attack on reference with the argument against reference set forth by Kripke in his interpretation of Wittgenstein on rule following.[38] Although the argumentative strategies of Putnam and Kripke are superficially similar, I believe there is an important difference—a difference that makes Kripke's argument by far the more formidable. (I shall refer to the argument as Kripke's, even though he says he is only expounding "Wittgenstein's argument as it struck Kripke" (*ibid.*, p. 5).)

---

[35] Ayer, p. 124 "[N]o individual sign refers beyond itself. . . [I]ts having a reference to such and such an object simply consists in the fact that it figures or at least is capable of figuring in a series of signs of such and such a sort [T]he series of actual signs must be finite. . . It is their membership of the series that gives the individual signs their meaning, rather than any special meaningful relation in which they stand to one another."

[36] It is further arguable that the price we pay is a reductive account of the objects of reference. Here Kant comes out looking better than Ayer's Peirce A phenomenalist theory of objects is perhaps not out of the question, but a verbalist theory of objects certainly would be!

[37] Putnam does not present himself as a nihilist, of course; his "internal realism" is supposed to save the day. He sometimes gives the impression that the rescue is to be effected by letting the referents of our terms be not the mind- and theory-independent objects of the metaphysical realist, but objects of a closer-to-home variety that exist somehow "within" our scheme (see RTH, 52). I think it is plain, however, that the difficulties Putnam raises for reference have to do with the relation itself and not with its relata If the difficulties are at all genuine, they make reference indeterminate or nonexistent for any theory, internal realism not excepted I say more about this in the paper cited in fn. 33.

Kripke's conclusion, as I take Putnam's to be, is nihilistic: there is never a fact of my meaning one thing rather than another by any word—e.g., no fact of my using 'plus' to refer to addition rather than *quaddition*, an operation that normally coincides with addition, but yields discrepant values at certain points. Here is how Kripke puts it:

> There can be no fact as to what I mean by 'plus', or any other word at any time. . . . [T]here is no fact about me that distinguishes between my meaning a definite function by '+' . . . and my meaning nothing at all (*ibid.*, p. 21).

In some passages, it appears that Kripke's core argument can be developed along the following lines. (See *ibid.*, pp. 15–22.) How can a teacher convey unambiguously to a learner that 'plus' refers to addition rather than some other operation? Anything the teacher *says* can be misinterpreted; and anything the teacher *shows* by means of examples will be compatible with many different extrapolations. Now let teacher and learner be contracted into the same individual: imagine that I am giving myself instructions to fix what I shall mean by 'plus'. If any instructions I give to a pupil leave the meaning of 'plus' unsettled, how can I do better with the instructions I give myself? How can any instructions regarding the use of 'plus' pick out one function rather than another that diverges from it at a certain point? Giving myself a finite number of examples that instantiate the intended function clearly will not suffice; and if I try to internalize a rule that determines the output of the function for every pair of numbers in its infinite domain, I only invite the question how the rule itself (or the symbols for it in my language of thought) are to be interpreted. For example, if the rule is stated in terms of counting, I invite the question how it is that by 'counting' I did not mean *quounting*, an operation that relates to quaddition as counting relates to addition.

It is at this point that it may begin to appear that Kripke's argument is similar to Putnam's—and similarly in error. Kripke coins

---

[38] *Wittgenstein on Rules and Private Language* (Cambridge: Harvard, 1982)

'qu'-words as freely as Putnam attaches asterisks; he repeatedly asks how it is that a word has one meaning rather than that of its quounterpart. It may seem that Putnam and Kripke are both querying the interpretation of items that should not be up for interpretation—of facts that are simply facts, not items that need stand at either end of a semantic relation.

In fact, however, I believe that Kripke's argument is apposite in a way that Putnam's is not. Kripke does not irrelevantly ask, as Putnam does, how we can *single out* the facts put forth as those on which semantic facts supervene. He does not query the interpretation of the sentences we use in an effort to state the facts belonging to the supervenience base. Instead, he challenges the adequacy of the facts as stated to determine semantic phenomena. When he asks Putnamesque questions about quounting and the like, he asks them about thought symbols, which are the very things (along with words) whose referring power is in question.

Kripke's overall argument has the form I set out in section I: it combines a premise of supervenience with a premise of underdetermination. His defense of the underdetermination premise proceeds by criticizing the sufficiency of various candidate facts in which meaning facts might be supposed to consist—e.g., facts about my qualitative mental history, facts about how I have used an expression in the past, and facts about how I am disposed to use it in new cases.[39] His defense of the supervenience premise is much slighter, but it is nonetheless there; it comes out in his rejection of the idea that meaning facts could be *sui generis* (*op. cit.*, pp. 51–3). Kripke's argument is thus built on semantic supervenience as an essential premise; Putnam's argument, by contrast, is undermined by it.

JAMES VAN CLEVE

Brown University

---

[39] It is also possible to discern in Kripke a general argument to the effect that semantic facts cannot flow from natural facts because semantic facts have a normativity that no natural facts could bestow. (See, e.g., the article by LePore and Loewer cited above.) Kripke urges this point chiefly against dispositional theories (see *op. cit.*, pp. 11, 37), but the point could be generalized against any naturalistic theory of meaning or reference. An additional premise would obviously have to be that the normative never supervenes on the natural.

BAYESIAN CONDITIONALIZATION RESOLVES POSITIVIST/REALIST DISPUTES

PARTICULAR positivist/realist disputes are a recurrent phenomenon in virtually all empirical sciences and scientists regularly shift their positions on such issues in the light of new observational evidence. Eventually, the weight of the evidence usually leads within the science in question to stable general consensus for or against any particular realist ontology that has been at issue. For example, such present general consensus has been reached in modern physics against the caloric hypothesis, in favor of the atomic theory, against ether theories, in favor of quarks, against local hidden-variable theories, in favor of quantized Lagrangian field theories, and so on. It is clear that, if we were able to set up convincing rational reconstructions of how the weight of evidence can logically induce such general agreements in science in one or other direction, such an achievement might go a long way toward clarifying the structure of the positivist/realist disputes that occur in general epistemology.

Recent philosophers of science, in spite of having been provided by mathematicians with an apparently completely general recipe for any rational effect of evidence on theory choice, namely, the rule of probabilistic inference known as Bayesian conditionalization, have generally regarded that recipe as inapplicable to positivist/realist disputes, and indeed have often gone further and supposed that any probabilistic account of scientists' beliefs in theories must necessarily systematically favor positivist positions. For example, that was Karl Popper's main argument for rejecting probabilism and opting for realism and improbabilism, and it is Bas van Fraassen's main argument for accepting positivism—he believes positivism is supported by probabilism, which he accepts.

In the following pages, I shall clear up this confusion by offering explicit Bayesian analyses of the rational weight of observational evidence in various positivist/realist disputes, thereby making it clear that the conversions from positivist to realist positions (and vice versa) which have occurred on so many occasions in the history of science have been warranted by valid probabilistic reasoning on the basis of observational evidence—notwithstanding the apparent theorems to the contrary which have been cited by philosophers of science. My examples will make it clear that such philosophers have conflated degrees of belief in quite different propositions.

In the following pages, I shall use the rule of Bayesian conditional-

0022-362X/92/8907/362-82

ization in the following form. Let H be a hypothesis or theory, let E be the observational evidence, let $p(H)$ be the subjective probability that the person in question ascribes to the theory H before the evidence E has come in, let $p(E, H)$ be the subjective probability that before he knows the truth of E, he ascribes to the truth of E conditional on the truth of H, and let $p(E)$ be the corresponding subjective probability that he initially ascribes to E.

$(p(E) = p(E, H)p(H) + p(E, -H)p(-H)$

(where $-H$ is the denial of the truth of H)

Then the subjective probability that he should rationally ascribe to H after he discovers the truth of the observational claim(s) E is, according to this standard rule of probabilistic conditionalization:

$$p(H, E) = p(E, H)p(H)/p(E)$$

In this formula, I shall substitute symbols representing various specific theoretical and observational claims for H and E, and I shall sometimes also need to subscript the ps in order to make it clear to which individual subjective probabilities, at some particular time, I am referring.

I. ABSTRACT GENERAL ANALYSIS OF POSITIVIST/REALIST DISPUTES

This section can be studied before or after the case studies that follow. In general, in each specific example of a positivist/realist dispute, I shall be concerned with four different substituents for H, namely, the detailed realist theory T_R, its observational consequence set T_P, the central realist ontological claim R, and the positivist denial of R, which I shall call P.

T_R will entail T_P, so $p(T_P)$ will always and for everyone be greater than or equal to $p(T_R)$. T_R will also entail R, so $p(R)$ will always and for everyone be greater than or equal to $p(T_R)$. P is simply the negation of R, so, for everyone, $p(P) = p(-R) = 1 - p(R)$. P (while always entailing $-T_R$) neither entails nor is entailed by T_P, however. Depending on the circumstances and the individual whose beliefs are at issue, $p(P)$ may be less than or equal to or greater than $p(T_P)$. This is not widely enough appreciated.

The philosophical conflation of P and T_P, and hence of $p(P)$ and $p(T_P)$, has in fact formed the greatest obstacle to any real understanding of what is going on in positivist/realist disputes. In fact, observational evidence often, as we shall later see, lowers $p(P)$ while raising $p(T_P)$, or lowers $p(T_P)$ while raising $p(P)$. Indeed, such scenarios prove to be the most typical ones.

This should have been obvious from the history of science, namely, that the more success a realist theory has in its observational predictions, the more scientists will be inclined to accept its central ontological claims, i.e., $p(T_P)$ rises and $p(P)$ falls, and the more it runs into trouble with its observational predictions, the more scientists will be inclined to call those central ontological claims in question, i.e., $p(T_P)$ falls and $p(P)$ rises. I shall show that precisely such changes are required by ordinary applications of Bayesian conditionalization.

In applying this formula, I shall freely ascribe degrees of belief/subjective probabilities to scientific hypotheses and theories. In general, such hypotheses and theories will not be decidable assertions, so they could not be directly betted upon. There is no more reason, however, why we should insist on simple operational definitions for the theoretical parameters in epistemic theories than we should for the theoretical parameters in physical theories. Indeed, it seems to me a subjective fact that the degrees of belief we assign to decidable statements are nearly always in practice introspectively derived from those we assign to more general less decidable statements. Furthermore, de Finetti's representation theorem shows (pace de Finetti) that assigning degrees of belief to decidable statements (e.g., outcomes of finite numbers of coin tosses) often commits us to assigning degrees of belief to undecidable statements (e.g., degrees of bias of coins).[1]

The only difficulty I recognize about assigning degrees of belief to scientific theories arises from the vagueness often inherent in such theories, e.g., where it is not specified in advance precisely what the intended scope of a particular such theory is, so that one can envisage possible worlds, or indeed future scientific scenarios, in which it would be largely a matter of future convention whether we decided that the original theory had turned out to be true or false. But such problems, when they arise, can always be solved by further initial specification of the theory.

My later worked examples will indeed presuppose that even the positivist assigns some small initial probability to R and to T_R. Here I would say that, even if the positivist claims on the basis of some theory of meaning that R is meaningless (actually this is practically never the case with real scientific examples, until philosophers start sticking their oars in), it would still be altogether unreasonable for the positivist to assign a probability of 1 to the correctness of his favored theory of meaning. (For he would be crazy to offer us infinite odds that he will not himself abandon that theory after more careful analysis of its implications.) So his $p(R)$ still should not come out as actually 0. In any case, anyone who wishes to do so could read all the negations in my later examples as indicating falsity-or-meaninglessness and the calculations would still go through.

---

[1] This philosophically important example was discussed with a proof in my "The Probabilities of Statistical Hypotheses," in *Applications of Inductive Logic*, L. J. Cohen & M. B. Hesse, eds. (New York: Oxford, 1980). Again, with Colin Howson, I regard as worthless, within a theory of personal probabilities, Popper's neo-Carnapian arguments to suggest that all scientific hypotheses should be assigned subjective probabilities of zero.

In any specific example, the central realist ontological thesis R itself is often initially rather vague and may become more precise as T_R is further articulated. This may lead to eventual identification of R and T_R, or, if T_R is itself developing internal paradoxes or running into observational difficulties, to a fairly sharp later separation between R and T_R. Where R itself remains vague it is often best to identify belief in R with a belief in the not-yet-exhausted·theoretical (and ultimately predictive) fruitfulness of a particular analogy. But in other cases a belief in R, which has become modified and qualified in the course of time as a result of further development of T_R, may itself eventually become indistinguishable from a belief in T_R.

As T_R is further articulated, T_P, the set of T_R's observational predictions, will in general grow. Also, T_R and hence T_P will usually be modified from time to time to improve the empirical fit of these predictions. The associated modifications in and further articulation of T_R, and correlated extensions of T_P, may lead to a progressive narrowing of the gap between T_R and T_P. That is, the initially obvious surplus content of T_R, relative to T_P, may disappear in the course of time, having been progressively cashed by modifications and refinements of T_R and correlated extensions of T_P. In such a case, a belief in R may finally become indistinguishable from, and logically reducible to, a belief that all actual and all possible observations would always satisfy a certain particular set of far-reaching mathematical constraints built into T_P. And the positivist rejection of R, and thus acceptance of P, would then become identical with skepticism as to the universal, timeless, and counterfactual-embracing validity of this particular set of far-reaching and inexorable mathematical constraints on all possible observations.

The disagreement between the realist and the positivist here would thus essentially reduce to an issue of how far one was entitled inductively to extrapolate from the hitherto actually performed observations. Thus, it would simply be a matter of different degrees of belief in T_P. But these differences might of course only arise in relation to parts of T_P's full content which were altogether remote from its practical content with respect to actual and technically feasible observations. In general, a T_P will include predictions of the observational outcomes along all branches of a vast tree of action choices, including all sorts of alternatives that are in practice mutually exclusive, and including specifications of what would have been the observational outcomes of unperformed possible past actions. It is only thus in a highly rarified sense that I shall here suggest that positivist/realist disputes are in an ultimate analysis reducible to differences of degrees of belief in sets of observational claims.

Furthermore, in these ultimate limiting cases T_P will typically be a theory on which we can obtain no finite grasp except via its T_R axiomatization. That is, we will know as mathematicians that all the theoretical terms we use are in principle eliminable, and actually observationally definable, and merely at the cost of a more complicated axiomatization. But this cost may be a completely prohibitive cost, not only for ourselves but possibly for any finite system. So, if we wish to believe in T_P, the only way to do this is to believe in T_R, and we will know from our mathematical metatheorems that this belief in T_R is in fact metaphysically innocent, we are not committing ourselves to anything not in principle reformulable in positivistic language. So the realist is really just a positivist who is happier about far-reaching observational extrapolations.

This is not the situation in lower-level positivist/realist disputes. At earlier stages in their history, realist theories T_R do have surplus content over their observational consequence sets, T_P. But this surplus gets cashed in terms of further observational consequences as the realist theory is further articulated. The positivist is generally more skeptical than the realist concerning the empirical scope and accuracy of the current T_P, and he certainly does not expect further observational successes as a result of further articulation of T_R; but the realist does; or the positivist expects fewer such further observational successes than the realist does. These are the real scientific disputes between positivists and realists. They are not metaphysical disputes. Subsequent evidence favors one party or the other, and Bayesian rationality requires corresponding belief shifts.

II HISTORICAL EXAMPLES

Consider the case of early nineteenth-century realism concerning atoms and its early nineteenth-century positivist critics. Let the realist theory of atoms be T_R. Let the corresponding positivistic observational theory, in this case the laws of *constant and multiple proportions*, be T_P. We note that T_P, in this as in all cases, goes well beyond the known experimental data. In fact, in this case T_P is a very far-reaching extrapolation from fragmentary and in some respects conflicting known experimental data. T_R entails T_P but T_P does not entail T_R. T_P could be true and T_R false.

For example, although T_P, if it turns out to be true without exception, might be regarded as entailing some funny kind of quantization rule governing the proportions of elements in compounds, that rule alone would certainly not entail any computable limit on the divisibility of matter of a given sort. Nor even if such a limit were computable, on the basis of further observational evidence and an extension of the theory, would T_P entail that the corresponding hypothetical smallest constituents of given macroscopic materials could be consistently localized relative to one another in ordinary three-dimensional space. Relative location in three-dimensional space might prove to be not even well-definable at the submicroscopic level in question. Nor according to T_P need it be the case that we could legitimately ask whether such hypothetical ultimate constituents were moving, and if so how fast, and so on. T_R, on the other hand, would be false, even if T_P remained true, if the world were such that future observations would eventually generate irresolvable inconsistencies with the assumptions implicit in T_R that atoms have finite sizes, have locations relative to each other in space, and so on. Thus, while there may be a certain amount of vagueness in what T_R precisely entails, it certainly embodies inter alia an empirically stronger constraint on future observations than does T_P as here identified. Thus, at later stages in the development of atomic theory we shall need a new T_P that asserts more than the laws of constant and multiple proportions. For the moment, though, I shall restrict myself to the original version of T_P.

For what follows, it is crucial to appreciate the following logical relations. T_R, the realist theory, entails R, the realist ontological position. R may also entail T_R, as is plausible in this historical example, so it does no harm provisionally to identify R and T_R. On the other hand, P, the positivist ontological position, does not entail T_P, nor does it entail the negation of T_P: to reject the existence of atoms one is committed neither to accepting nor to denying the validity of the laws of constant and multiple proportions. P is thus simply $-R$, and hence in this case, provisionally, identical with $-T_R$. (Any philosopher who has qualms here as to whether positivists merely assert the falsity of realist ontological claims or actually claim they are meaningless can simply read the symbol '$-$' as 'false or meaningless' without affecting the validity of the following arguments.)

On the other hand, T_R does entail T_P. Indeed, even independently of our possibly over-simplified identification of R with T_R in this historical example, R itself here pretty clearly entails T_P. So in any case it follows (from a theorem of elementary probability theory) that T_P is at least as probable as, or more probable than, T_R. But nothing follows from that about the relative (subjective) probabilities of P and R, or of P and T_R. In particular, T_R can still be more probable than $-T_R$, hence, a fortiori, R more probable than P. Thus, the Popper/van Fraassen inference from the relation between $p(T_R)$ and $p(T_P)$ to the conclusion that positivist ontological claims have always to be as or more probable than realist ontological claims is of course nonsense.

Now, let us subscript our *p*s according to whether they represent the degrees of belief of the realists, $p_R(\,)$, or those of their positivist opponents, $p_P(\,)$. Let the realist at a given point in time assign a subjective probability of 0.6 to T_R: $p_R(T_R) = .6$, that is, the realist thinks that it is a little more likely than not that Daltonian atoms really exist. Hence,

$$p_R(T_P) = p_R(T_P, T_R)p(T_R) + p_R(T_P, -T_R)p(-T_R)$$

according to a standard identity of the probability calculus. Hence, $p_R(T_P) = 1x.6 + p_R(T_P, -T_R)x.4$. What is the realist's estimate of the only unknown here, namely, the subjective probability that he would assign to the truth of the laws of constant and multiple proportions were he to believe that Daltonian atoms did not really exist, thus merely on the basis of the direct observational evidence so far available? Let us for the purpose of discussion take this as 0.2. It would follow that $p_R(T_P) = 1x.6 + .2x.4 = .68$. As Popper and van Fraassen and others have observed, the realist thus ascribes here a higher probability to the positivist version of the theory than to the realist version. Nevertheless, the realist here trivially assigns a higher probability to T_R than to $-T_R$ and hence, a fortiori, to R than to $-R$, hence to R than to P.

What of the positivist? Suppose at the same point in time he ascribes a subjective probability of only 0.2 to the existence of Daltonian atoms. Then $p_P(T_P) = 1x.2 + .2x.8 = .36$. Notice that the positivist and the realist are here assumed to agree on how probable the laws of constant and multiple proportions would be merely on the basis of the direct observational evidence. In this, though not in all, positivist/realist disputes, this assumption is historically plausible. Notice also that .36 is smaller than .68, thus, given these particular conjectural historical subjective probabilities, the positivist is not only less sure than the realist of the exact truth of the observational laws, namely, the laws of constant and multiple proportions, but in this particular case he thinks that they are more probably than not false (for example, that future evidence will produce counterexamples). This is again historically plausible for this example.

What happens when further observational evidence for the truth of T_P is obtained? Well, the maximal further observational evidence that could be obtained would be observational evidence that brought the subjective probability of T_P arbitrarily close to 1. Let us see what effect this would have on the degrees of belief of the realist and the positivist. What we need to calculate is thus, by Bayesian conditionalization,

$$p_R(T_R, T_P) = p_R(T_P, T_R)p_R(T_R)/p_R(T_P) = 1x.6/.68 = \text{roughly } .9$$

and

$$p_P(T_R, T_P) = p_P(T_P, T_R)p_P(T_P)/p_P(T_P) = 1x.2/.36 = \text{roughly } 6$$

In other words, the realist is merely strengthened in his realist belief, but the positivist has actually converted to atomic realism.

The reader is invited to rework these calculations for himself with other numerical inputs. He will find for example that, if the positivist had ascribed an initial subjective probability of only .1 to the truth of realism but .2 (as in our example) to the initial probability of the laws of constant and multiple proportions purely on the basis of the observational evidence, then he would never actually convert to realism purely on the basis of further observational evidence for those particular observational laws, though he would come close to it. On the other hand, if he had assigned a higher initial "purely observational probability," say, .6, to those laws, he would need to have started with $p_P(T_R)$ close to .4 to have eventually converted to realism concerning atoms on the basis of further observational evidence for the truth of those observational laws.

The latter numerical example has also the interest that it illustrates a situation where the positivist does, even initially, regard the observational laws as more likely to be true than not (though not as likely as the realist considers them) and he still will convert eventually to realism, given sufficient further evidence for them. This shows that the accident in my original example that the positivist began by thinking the relevant observational laws were more likely than not false (which I regard as a plausible reconstruction of the actual historical situation with respect to this positivist/realist issue) is by no means essential to the possibility of subsequent rational conversion to a realist position.

So far I have treated the laws of constant and multiple proportions as "purely observational." That is of course an oversimplification. "Observationality" is always relative to the particular positivist/realist dispute at issue. A more extreme positivist could certainly detect theory ladenness in these laws, and in relation to him we would have to reanalyse the situation with respect to a yet more positivistic observation language, and the corresponding "observational" laws as formulated within that more restricted observational language.

It might seem from this example that the direction of conversion is always eventually from positivism to realism. That of course is not true. Not only would a refutation of the observational laws derivable from the realist theory refute that realist theory, leaving the positivist position intact, but also observational evidence could accrue which would diminish the probability of those observational laws without actually refuting them. Indeed, this is what happened for a period after Dalton's original publications: the observational laws in fact ran into trouble with alloys and this trouble was not cleared up until alloys were clearly distinguished from both compounds and mixtures. So for a time $p(T_P)$ decreased for everyone, and probably enough—a good approximate calculation of the rational opinion change in question can be constructed simply by applying Jeffrey-conditionalization on a lowered value of $p_R(T_P)$—for some realists concerning atoms briefly to reconvert to the positivist position.

There also arose further observational difficulties for the realist position. Gay-Lussac's observational laws on combining volumes of gases seemed, when conjoined with the atomic theory, to require the Avogadro-Ampère hypothesis. But that hypothesis seemed at the time very implausible. Quite apart from the difficulty of understanding the required polyatomic rather than monatomic molecules of even elementary gases, the existing static atomic theories of temperature and pressure of gases failed to render the Avogadro-Ampère hypothesis intrinsically plausible; and such incipient kinetic theories of gas pressure as were then available suggested rather that gas pressure at a given temperature ought to increase directly with the masses of the molecules, because this would, for given speeds of molecules, increase the momentum transfer, and hence pressure, at the boundaries. Formally, with A as hypothesis, and G as Gay-Lussac's observational laws, the crucial issue was whether $p(A, T_R)$ was not perhaps even smaller than $p(G, -T_R)$, where the latter corresponded roughly to the estimated chance, independently of the atomic theory, of there being a further mysterious quantization rule corresponding to Gay-Lussac's findings. At the time many atomists appear to have estimated the value of $p(A, T_R)$ as so small that they were more inclined to regard Gay-Lussac's original results as an experimental artifact than to welcome them as additional further support for T_R.

Of course, nearly fifty years later, with the development of the modern kinetic theory, $p(A, T_R)$ was drastically revised upward, and as a result, after the clarification of the chemical situation by Cannizarro, $p(G, T_R)$ became much larger than $p(G, -T_R)$, hence G now provided, precisely in accordance with Bayesian conditionalization, further strong evidence favoring realism concerning atoms.

The later predictive successes of structural atomic chemistry, and the increasing agreement between estimates of the size of atoms obtained from very different calculations based on very diverse classes of phenomena, provided further evidence E for which $p(E, T_R)$ was very much greater than $p(E, -T_R)$, and hence, by Bayesian conditionalization, further reason to accept realism concerning atoms.

In the last quarter of the nineteenth century, the remaining doubts concerned not so much whether matter took on a discrete structure at very small dimensions—even Ostwald seems to have accepted that already in the 1880s—but rather whether the vast velocities imposed by the kinetic theory on these smallest discrete parts were established beyond reasonable doubt by the observational predictions to which they led. There were at least three difficulties here: (i) some of the predictions (e.g., concerning specific heats) seemed either wrong or to require fantastic hypotheses concerning the dynamical structure of molecules; (ii) in the case of the major successful predictions of the molecular-kinetic theory, there was a strong suspicion—supported by authorities such as Ostwald, Duhem, and Planck—that most of the successful predictions could be rederived on the basis of purely thermodynamical considerations, hence that for these predictions $p(E, -T_R)$ was not substantially different from $p(E, T_R)$, where T_R is now the fully extended molecular-kinetic theory; and (iii) early attempts to explain Brownian motion on the basis of the kinetic theory seemed to give quantitatively completely wrong results, so the evidence of Brownian motion seemed at the time to count against rather than in favor of the kinetic theory. With B as the quantitative observational features of Brownian motion, it seemed that $p(B, T_R)$ was substantially less than $p(B, -T_R)$, hence B seemed to be rational probabilistic evidence against T_R.

Problem (i) was cleared up in the early twentieth century with the development of the old quantum theory; (ii) was clarified as it gradually became clear that the energeticists and thermodynamicists had been overly optimistic concerning how many of the predictive achievements of the kinetic-molecular theory they could in fact reproduce without adding arbitrary extra assumptions to their phenomenological theory; and (iii) was finally resolved by the theoretical work of Einstein and Smoluchowski and the experimental work of Perrin. The quantitative details of Brownian motion, B, thus eventually became a further piece of evidence for which $p(B, T_R)$ was finally recognized as close to unity and $p(B, -T_R)$ as very small indeed: thus, B constituted observational evidence with enormous evidential weight according to Bayesian conditionalization. It was indeed sufficient to convert the earlier positivistic-skeptic Ostwald to the molecular-kinetic theory.

We note that in the course of these developments, T_P, the positivistic reconstruction of the predictive content of the atomic theory, has to be constantly enlarged, and indeed has to have more and more quantization features built into it. At the same time, T_R, the realist theory of atoms, becomes refined (especially with the development of the quantum theory of the structure of matter) in such a way that atoms lose many of the features originally attributed to them on the basis of a naive analogy with ordinary macroscopic particles. The question in fact arises as to whether this progressive strengthening of T_P (which of course by the later twentieth century will have to include in its predictive content the results of x-ray crystallography, of electron microscopy, of "the making visible of individual atoms" using scattered laser light, and so on) and modification of T_R in such a way that its details become more and more tied to details of T_P, does not eventually lead to formal mathematical equivalence between these two theories.

Here we must never forget that T_P is a predictive observational theory that must contain among its predictions not only all the actually confirmed observational predictions of T_R, but also all its actual and conditional future predictions, including expectations conditional on the development of future, as yet unthought-of instrumental techniques for the further probing of the hypothetical submicroscopic reality. Could it not be that the modern T_R is simply a more economical expression of the modern T_P?

That there may no longer be a genuine logical gap here between the realist and the equivalent positivist theory is further suggested by the position of the only surviving antiatomist I have come across, namely, C.A. Truesdell. Truesdell's position—expressed in conversation, I do not know whether it is in print—is that the best modern atomic theory we have is in fact a quantized field theory of matter in which the number of atoms is merely a quantum number, and that the nearest classical equivalent to such a theory is not a discrete particle theory of matter but a continuous field theory of matter. Hence Truesdell maintains that it is really the sophisticated antiatomists rather than the naive atomists who have been subsequently vindicated. Whatever the merits of this reading of history, it suggests that antiatomists and atomists now believe in essentially the same theory and that the difference between them has become at most terminological: T_P has been extended so as to include most of the extra content of the original T_R, and other initial expectations of T_R have been so qualified that the two positions seem no longer clearly distinguishable.

Surprising though it may seem, there are in fact a few rather clear examples of positivist/realist disputes where T_P seems not only to be entailed by, but also to entail, T_R. For example, in Mach's reaxiomatization of classical Newtonian mechanics, although masses are eliminated from his positivistic axiom system—which speaks instead of inverse ratios of mutually induced accelerations, which ratios are posited to be the same whether mediately or immediately arrived at—nevertheless the existence of masses, unique up to a choice of unit, is then trivially provable within his axiom system. (Patrick Suppes' later axiomatization failed to yield this result because it assumed a funny observational language; see fn. 2.) The difference is that 'mass' does not occur as a primitive term in Mach's positivistic formulation of the theory. While Mach's T_P is thus actually formally equivalent to Newton's T_R here, the positivist ought to be less certain of the truth of either than is the realist. In particular, the positivist here can, and I would say must, take very seriously the possibility that the fact that the ratios of accelerations in question turn out the same whether mediately or immediately arrived at may turn out to be true merely as some kind of first-order approximation, or to possess a class of exceptions: in which case realism concerning masses would be false: for the observational laws would fail to satisfy the mathematical conditions required for the existence of the mass functions. Thus, in this case the choice of axiomatization is related to the theorist's suspicions as to where the theory is likely to break down.

The realist over masses need not be absolutely confident that the orthodox theory of mechanics will never require future modification, but he does not expect it to require future modification of such a kind that masses will cease to be well-defined parameters; or, at the very least, the positivist considers this a much more likely eventuality than the realist does. Hence, at least in this respect, we expect $p_P(T_R) = p_P(T_P)$ to be substantially less than $p_R(T_R) = p_R(T_P)$. It then follows by Bayesian conditionalization that observational evidence—e.g., further and further evidence that T_P (or the relevant part of T_P) really is exactly correct and not merely a good first-order approximation in our actual universe—will shift the positivist's beliefs in the direction of realism. For T_P, or the relevant part of it, is itself here construable as observational evidence O such that $p(O, T_R)$ is greater than $p(O, -T_R)$.

It is historically likely that Mach's positivism concerning masses is of a less strong sort than the positivism (shared by Mach) that we have previously considered concerning atoms. That is, it seems to me plausible here to reconstruct Mach's position as assigning to T_P and T_R here a subjective probability of, say, 0.6, compared with the realists' assignment of, say, 0.99. The difference is one of degree of confidence, Mach's degree of confidence not being such that he is willing to choose a formulation of the theory which cannot easily be modified should the observational laws required for the mathematical existence of masses run into experimental difficulties. The transition from Mach's relative positivism to relative realism in this case will probably thus occur not at the .5 confidence level but at some vague higher point, say, around .9, at which it seems to become reasonable no longer to allow, within the formalism, for a rather unlikely kind of possible future modification. Clearly, increasing observational confidence in the observational theory T_P will directly shift Mach's position here, since $p(T_R)$ in this case equals $p(T_P)$.[2]

There are in fact other examples of the same relation between positivist and realist positions. For example, Poincaré once reaxiomatized Newtonian gravitational theory taking relative rather than absolute accelerations as a primitive. The axiom system required turned out slightly more complicated than the usual one, but of course one could prove within it that there existed a set of reference frames such that accelerations relative to that set of frames satisfied precisely the usual axioms governing absolute accelerations. In other words, in this case too, T_P turned out not merely to be entailed by, but also to entail T_R. In fact, Poincaré seems to have been satisfied enough of the truth of the relevant observational content of the orthodox realist theory here not to argue that his more positivistic reaxiomatization yielded a preferred choice of axiom system.[3]

---

[2] Readers who have been persuaded by the arguments of Suppes and others that Machian-style reaxiomatizations fail to eliminate the theoretical entities they are supposed to eliminate are referred to my "The Eliminability of Masses and Forces in Newtonian Particle Mechanics: Suppes Reconsidered," in the *British Journal of the Philosophy of Science*, XXVIII, 1 (1977): 55–7. One can add to the objections presented in detail there, the objection that simple Padoa's principle arguments do not go through in our actual universe because it lacks the necessary perfect symmetries exploited in those arguments. Sneed's conjectures concerning Ramsey-noneliminability also do not in my view stand up to careful technical scrutiny The best neo-Machian positivistic axiomatization of Newtonian mechanics is to be found in the first part of G W. Mackey's *Mathematical Foundations of Quantum Mechanics* (New York: Benjamin, 1963), and its only slight awkwardness in ending up with masses which could be different in different directions is easily avoided by imposing rotational invariance on any laws from the beginning. The observational definability of ordinary theoretical terms is entirely clear in this presentation

[3] Further examples of positivistic reaxiomatizations of standard physical theories, where the existence and uniqueness of the eliminated theoretical entities is rederived within the positivistic re-axiomatization, can be found in Robin Giles' studies of the mathematical foundations of thermodynamics and of quantum theory.

A further instructive example of the same kind is the relation between Einstein's original positivistic formulation of special relativity theory based on the Lorentz transformation equations and the realist formulation of that theory in terms of Minkowski space-time geometry. The entities taken to represent underlying physical reality in the space-time geometrical formulation are simply the invariants of the transformation group relating reference systems, i.e., of the transformation group that is taken as fundamental in the more positivistic formulation of the theory. But clearly the truth of these transformation laws mathematically entails the existence of the space-time geometrical invariants in question. So here again T_P entails T_R as well as vice versa.

Those more positivistic theorists who regard the transformation equations as more fundamental than the corresponding space-time geometrical formulation (a formulation in which observer's-frame-of-reference dependence is no longer a feature of any of the primitive concepts) probably do so because they envisage a future in which those transformation equations will be seen as merely part of a wider group of transformations to which the dynamical laws of our universe are subject and which will require a deeper theory than space-time geometry for its explanation. Such a theory would eventually set limits on the validity of space-time geometry and of those transformations, and perhaps also on the observer-independence of the properties and concepts we need to take as fundamental. Realists concerning Minkowski space-time geometry, on the other hand, are more confident that essentially geometrical explanations of the Lorentz transformations will remain fundamental in future theory, and refuse to regard special relativity theory as providing any evidence for expecting any ineliminable observer-dependence as a feature of future fundamental theories in physics. It seems thus that while both groups of theorists will already admit that, given the success of general relativity, both T_P and T_R must be regarded here as of only local validity, the positivists are less confident of both of these, for they suspect that T_P is only a small part, valid to a high approximation, of a future more global theory. Hence the situation here is rather similar to the situation with respect to Mach's positivism concerning masses in classical mechanics.

These examples show that it is possible to make sense of positivist-realist disputes even in cases where the realist theory is in fact logically and mathematically formally equivalent to the positivist theory. The issue here between the positivists and realists is in fact a matter of degree of confidence in the theory in question. We can identify this theory indifferently with T_P or T_R since here these mutually entail each other, so for both realist and positivist, $p(T_P) = p(T_R)$. It is $p_P(T_P)$ (= $p_P(T_R)$) which is less than $p_R(T_P)$ (= $p_R(T_R)$). Contrary to what one might expect, $p_P(T_P)$ is not greater than $p_P(T_R)$ nor is $p_R(T_R)$ greater than $p_R(T_P)$.

The question therefore seriously arises in the light of these examples whether it may not also be in the case of modern atomism that, as a result of convergence in the content of T_P and T_R, we have reached a point in which the same mutual mathematical entailment holds as in the cases of Machian (and neo-Machian) and Newtonian formulations of classical mechanics, Poincaréan and Newtonian formulations of classical gravitation, and Einsteinian and Minkowskian formulations of special relativity.

Precisely the same question arises as to the relations between phenomenalism and realism concerning the external world. May not a full phenomenalist T_P, were we able to write it down, not merely be entailed by, but logically and mathematically actually entail, the corresponding T_P? The logical relationship could be like that between a large table of correlation coefficients and the small number of factors resulting from a complete factor analysis of that table. Clearly this latter example shows that a mere mathematically more economical redescription of the data can be viewed as having considerable explanatory value, even though this is neither a Hempelian nor a Popperian explanation, since the explanans has no greater scope than the explanandum.

But there is an even more direct mathematical analogy between the epistemological phenomenalism-realism issue and the Einsteinian and Minkowskian formulations of special relativity. In both cases we seem, on the one hand, to have a phenomenological theory that describes and relates all possible perspectives and, on the other hand, a realist theory that posits an objective multidimensional geometry that is in fact built up precisely out of the invariants mathematically generated by the corresponding perspectival theory. Formally, one expects the two theories to be equivalent and mathematically to entail one another. This is perhaps not obvious in the case of phenomenalism relative to the external world or positivism with respect to modern atomic theory merely because no one has taken the trouble formally to construct the T_Ps in question, for they would so obviously be clumsy formulations compared with the corresponding realist formulations. But there is every reason to suppose that the set of all possible perspectival views of a group of objects (whether or not supplemented by a topology associated with possible trajectories of observers through this set) would possess a natural topology that was three-dimensional and that the ordinary realist description of the shapes and sizes and relations of those objects in three-dimensional space would simply be derivable as the mathematical structure of the mathematical invariants underlying the corresponding perspectival group of transformations. Ordinary realist claims would thus be recovered as mathematical theorems within this phenomenalist description. (And extension of this phenomenalist data set to include proprioceptive sensations would not change the logical situation, but merely generate multiple derivability of the realist theorems.)

If these extrapolations from mathematically better understood situations are correct, the difference between a positivist and a realist in general epistemology is not so much a matter of their having different expectations as in their differing in the confidence with which they hold these expectations. This confidence can be raised or lowered, we have seen, in the ordinary way on the basis of favorable or unfavorable observational evidence, by the rule of Bayesian conditionalization. Thus, positivist/realist disputes do not differ essentially in their logical structure from any other more obviously purely empirical disputes in science. Bishop Berkeley is simply less confident than the realist that his re-entering-the-study-like experiences will always be followed by as-if-perceiving-his-desk appearances.

The positivist in any particular dispute may have perfectly reputable reasons for having less trust than the realist in the very far-reaching observational extrapolations from the actual observational data required by that particular realist position. For example, the phenomenalist may have little confidence in the ultimate success of physiological explanations of dreaming and of mystical experiences and, for reasons of his own, may consider an ultimate mentalistic ontology more plausible than an ultimate physicalistic ontology: he is thus less committed than the realist to the view that all future observational data will conform to a set of laws whose invariants are those required by common-sense physical realism. The success of microrealist theories of a macrorealist world constitutes of course further evidence in favor of macrorealism, for the additional law-like constraints that such success imposes on the observational data can only be regarded as having arisen by mathematical accident from the point of view of the phenomenalist or solipsist. Hence $p(E, R)$ is greater than $p(E, P)$ for this evidence E. Berkeley had in fact doubts about the internal consistency of the microphysical theories of his day, which doubts enabled him to attach less weight to this argument than it might otherwise have carried.

While Berkeley's doubts concerning the microrealist physics of his day have been more or less resolved by subsequent developments, new doubts have arisen in the present century with respect to our latest fundamental theories. Realism requires that all true observational laws be consistent with and explicable on the basis of an observer-independent reality. This imposes a quite strong mathematical constraint on those laws. The laws of the quantum theory apparently fail to satisfy this constraint, as can be seen from a consideration of either the no-hidden-variable-theorems or the quantum-mechanical measurement paradox. Prima facie, such a situation thus provides observational evidence against an observer-independent reality, hence against realism and favoring phenomenalism or (taking into account the problems involved in giving a relativistically-invariant, observer's-frame-of-reference-independent account of wave-packet collapse in measurement) even solipsism. There are a number of theoretical options, however, none of them very plausible, which remain open to the microrealist as a possible escape from the relevant no-go theorems. So, while the evidence from the strange observational laws embodied in the quantum theory weakens the case for realism—for these observational laws O, $p(O, R)$ is apparently much less than $p(O, P)$—and while some theoretical physicists have apparently converted on the strength of this evidence to versions of phenomenalism or solipsism, many others remain unconvinced, presumably because their prior confidence in realism was initially higher, and they still hope to find an acceptable interpretation of the quantum theory which frees it from commitment to a fundamentally observer-dependent reality. For them, $p(R, O)$ still remains greater than $p(-R, O)$.

I have argued so far as if the realist position, R, can always be identified with the realist theory T_R, although the positivist position, P, cannot generally be identified with the positivistic theory T_P. This simplification obviously requires certain qualifications. In the quantum mechanical case, we do not have any convincing T_R available: the realist merely hopes that such a theory will become available in the future. And even in the case of atomism, we saw earlier how T_R became changed and extended in the course of time, incorporating the kinetic theory and later specific theories of atomic structure; so that, if we restrict R here merely to realism concerning the existence of atoms, R clearly can no longer be directly identified with T_R. Indeed, it would already obviously have been an oversimplification to identify R with a belief in every detail of Dalton's original T_R. So in these cases, while R does not really entail T_R, T_R still entails R, so if T_R becomes subjectively more probable than not by Bayesian conditionalization, a fortiori so does R.

There are also cases where, surprisingly, T_R is either unavailable, or, where it exists, is less probable than not, given the truth of R, and where, contrary to the situation in our original examples, certain evidence for T_P counts against rather than for the truth of R. Quantum-mechanical realism, where T_R is usually at best programatic, may be an example of this, since at least some realist programs here actually require that some of the observational predictions of the standard quantum theory be false. For example, the experimental tests of the Bell inequalities excluded certain realist options that previously had their defenders. This situation where R has to be sharply distinguished from T_R arises acutely in situations where the realist theory T_R is a makeshift devised to reconcile certain originally unexpected aspects of a successful observational theory T_P with the ontology required by R.

A simpler example than quantum-theoretical realism is the neo-Lorentzian ether-theoretic realist interpretation of special relativity theory. Here, T_R is indeed set up so as to yield the observational equivalence of all inertial reference frames demanded by Einstein's T_P. But R itself, realism with respect to the ether, requires the existence of a theoretically preferred frame, and is thus inconsistent with a belief in the universal validity of the observational laws embodied in T_P, and hence, a fortiori, requires that the present T_R, which is set up so as to yield precisely T_P, can only be some sort of approximately valid theory. So in this case more and more observational evidence for T_P, and hence for T_R, actually favors the positivist's rather than the realist's theoretical position with respect to the ether This is the opposite situation to the one we met with in the case of the atomic theory and in the case of ordinary common-sense realism. Formally the situation is easy to understand, however. If O is new observational evidence casting doubt on the equivalence of all inertial frames, $p(O, T_R) = p(O, T_P)$ is very small and approximately equal to $p(O, P) = p(O, -R)$, whereas $p(O, R)$ is substantially larger; hence $p(R, O)$ is greater than $p(P, O)$, and Bayesian conditionalization would here yield a shift of opinions in the direction favoring the ether-theoretic realists. Conversely, further evidence favoring the observational equivalence of all inertial frames—for example, new evidence discrediting that particular aforementioned O or its theoretical interpretation—would cause a rational opinion shift in the direction of the positivist position with respect to the neo-Lorentzian ether.

Similar situations arise with respect to a few other positivist/realist disputes in physics: for example, one form of realism concerning vector potentials in electromagnetism would insist not only on their mathematical existence (which is provable within the corresponding positivist theory) but on their uniqueness modulo a smaller class of transformations than the full class of gauge transformations. But, like the uniqueness of the neo-Lorentzian ether frame, this ontological conjecture has to be stated as an expected mathematical property of some future modified or extended observational theory: it cannot simply be imposed on the current T_R, since the T_P which that entails yields uniqueness only modulo the full class of gauge transformations: one could thus not impose further mathematical uniqueness within the current T_R without generating a mathematical contradiction within that T_R.

A failure to distinguish P ($= -R$) from T_P has confused most philosophical discussions of positivism and realism. But in these last examples, identification of P and T_P would be innocuous; philosophical confusion could easily be produced, however, by a failure to distinguish R ($= -P$) and T_R.

Attempts have been made in some of the physics literature to illuminate the peculiar features of some positivist/realist disputes by attempting to draw a distinction between the purely mathematical existence and real physical existence of the disputed entities or to reconcile their mathematical nonuniqueness with their supposed physical uniqueness. In my view, this only leads to obscurantism; for all these disputes can be fully clarified without introducing such doubtful, and probably ultimately unintelligible, distinctions; it is enough to take into account theorists' full future-observational expectations and the degrees of confidence they associate with these.

The same kinds of formal and evidential considerations apply, I believe, to all positivist/realist disputes in science and epistemology. It makes little ultimate difference whether the realist ontology in dispute is initially introduced on the basis of an analogy, as in the case of atoms, the ether, or other minds, or merely on the basis of a mathematically more economical reformulation of the observational theory based on exploitation of the higher-order-invariants structure of the original observational laws, as in the case of forces, potentials, wave functions, three-dimensional spatial geometry, or (3 + 1)-dimensional space-time geometry.

In the former cases, T_R is initially certainly stronger than the initial set of observational laws T_P that T_R was introduced to explain. But this distinction can be expected to evaporate in the course of time with the introduction of empirical qualifications on the original analogy and extensions of T_P to incorporate and render precise all the more vague and indirect observational expectations implicit in the original T_R. For example, even in the case of the other-minds issue, T_R must allow for empirical differences between my mind and those of others, and T_P must be extended to include physiological as well as behavioral evidence, even including the possibility of future electrophysiological gadgetry allowing rather direct access to another's thoughts and feelings. And here the ultimate success of a detailed functionalist theory of the mind rendering T_P and T_R ultimately fully equivalent cannot be altogether discounted. Even if this is not achieved, there remains no obvious bound on the posterior probability of the realist theory which can be reached here by successive Bayesian conditionalization on new behavioral and neurophysiological evidence. It is only with an unchangeable and unextendable T_P and an unqualifiable T_R that $p(T_R)/p(T_P)$ could remain fixed, and less than 1, in such a way as to set a bound on the maximum posterior probability of T_R. I can find no plausible examples of positivist/realist disputes involving such a constraint.

The widespread philosophical prejudice to the effect that realist conclusions can never be mathematically derived from positivistic axiom systems is not only open to numerous mathematical counterexamples, as we have already seen, but seems to me to be based on a false analogy going back at least as far as Plato. The analogy is that of the projection of the real three-dimensional world onto a single flat two-dimensional surface, the world of observations. But is it not obvious that, if we had such projections onto arbitrarily many different such surfaces, the three-dimensional reality might be uniquely mathematically reconstructable? A similar mathematical prejudgment was made in medieval astronomy when it was assumed by positivistic astronomers that astronomical observations could only determine the projections of the relative planetary positions onto the background sphere provided by the fixed stars, hence that, though the latitudes and longitudes of the planets were empirically meaningful, their altitudes (i.e., relative distances from the earth) were not. But here it only needed the realization that astronomical observations can be made from different points on the earth's surface for it to become geometrically clear that the planetary altitudes are in principle fully determinable (indeed, mathematically multiply overdetermined) by the parallax effects at the observational level. The full set of observational projections is more than sufficient to determine the actual locations of the planets in physical space demanded by the realist theory.

In the history of science there are many similar examples of realist conclusions turning out, surprisingly enough, to be logically derivable from positivist premises. A celebrated example is Heisenberg's matrix mechanics as contrasted with Schrödinger's wave mechanics. The former was explicitly set up as a strongly positivistic theory in which the only observables were essentially the frequencies, intensities, and polarizations of spectral lines. Schrödinger's theory, on the other hand, was set up as a realist theory describing the distribution of electronic mass and charge within the atom through its wave functions for stationary states. To Heisenberg's considerable surprise, the mathematical existence of the Schrödinger wave functions proved to be logically derivable within his own explicitly positivistic theory. (This example should not be confused with the later positivist-realist issues concerning the quantum theory. Heisenberg's original positivistic program that led to the setting up of matrix mechanics was much more extreme—more restricted in its observation language—than any later "Copenhagen" positivistic position.)

The more usual situation is, of course, where the realist theory comes first and only subsequently is more positivistically reaxiomatized. No mathematical physicist is then surprised when the existence, and uniqueness modulo the relevant groups of transformations, of the functions and parameters whose existence the realists merely posited proves mathematically derivable within the more positivistic axiom systems.

Realist ontologies, I thus suggest, probably ultimately simply posit the existence of mathematical objects (functions, invariants, and so on) whose existence would be actually mathematically required by a suitable set of positivistic observational laws, provided those full observational laws did indeed rigorously possess the exact mathematical structure required for the relevant existence and uniqueness theorems. Sufficient observational evidence will then (as the history of science teaches us) resolve, by ordinary Bayesian probabilistic inference, any particular positivist/realist dispute to any desired level of subjective confidence. I thus conclude that positivist/realist disputes are ultimately not metaphysical disputes at all but rather ultimately disputes as to the legitimacy of certain far-reaching observational extrapolations from the existing observational data, and that such disputes are thus resolvable (to any desired degree of convergence-of-opinion) by sufficient relevant observational evidence, in accordance with the ordinary rules of probabilistic reasoning.

JON DORLING

University of Amsterdam

BOOK REVIEWS

Dear Carnap, Dean Van: The Quine-Carnap Correspondence and Related Work. W. V. QUINE and RUDOLF CARNAP. Richard Creath, ed. xv + 484 p. Berkeley: University of California Press, 1991. Cloth $34.95.

W. V. Quine was twenty-four and Rudolf Carnap forty-one when they met in a hospital in Vienna at the end of 1932. That is perhaps the best age, and age difference, for mutually productive collaboration between teacher and pupil in philosophy. Ludwig Wittgenstein and Bertrand Russell were similarly related during their rich exchanges in Cambridge twenty years earlier. It might be interesting to see if other examples bear out the generalization. Quine followed Carnap back to Prague and had what he has described as his "first experience of sustained intellectual engagement with anyone of an older generation, let alone a great man." By then he had already studied at the Harvard of Sheffer and C. I. Lewis, and had written his Ph.D. thesis with A. N. Whitehead. In Prague he eagerly listened to Carnap's lectures on what was soon to become *The Logical Syntax of Language,* and on off days Carnap graciously invited him to his home for long discussions of the new work. Carrying the message back across the Atlantic, Quine gave three lectures at Harvard in 1934 to explain to a slightly wider world the basis of Carnap's conception of philosophy—or whatever was valid in what used to be philosophy—as syntax.

Those lectures have been part of the lore of Emerson Hall for many years, and it is very good now to see them in print. They appear here as the first fifty-odd pages of this thick volume, followed by the fruits of a correspondence between the two philosophers lasting almost thirty-eight years until Carnap's death in September 1970. The book concludes with Quine's "Homage to Rudolf Carnap," presented to the Philosophy of Science Association a month later. It is the only previously published material here. Richard Creath contributes a long introduction to explain to the uninitiated what the philosophical "battle" (as he calls it) between the two friends was all about.

These are not men to bare their souls to each other. There is obviously affection and great respect between them. They exchange expressions of genuine best wishes, warm regards, and occasionally love, with wives and children included. There are a few lively notes from Carnap's wife, Ina. The fortunes and whereabouts of the sev-

0022-362X/92/8907/383–6

eral Quine children through the years are warmly communicated. There is much news, especially on Quine's side, of travel. There are signs of academic maneuvering, especially in the earlier years when Quine is trying to get Carnap a job in the United States. His first choice is naturally Harvard, and he gives Carnap the names of "allies" among influential university figures to whom he should send "all future easily understood publications." Quine later asks Carnap to drop a word in the right places when he finds his own department's recommendation of his promotion rejected and thinks of leaving Harvard. Carnap obliges, but it turns out to be unnecessary.

There are mentions of health from time to time, almost exclusively physical, and mainly in connection with Carnap's chronic back troubles. Reports of meetings with mutual friends, mostly on Carnap's part, are largely restricted to what philosophical issues were discussed. A strong ideological solidarity which usually goes without saying surfaces in Quine's invocation of "the hard-headed, anti-mystical philosophical temper which all of us share" and in his reference at one point to "the enemy camp so to speak." The bonds are real, but remain largely unstated, or perhaps unexplored. Aside from a one-sentence announcement of his separation from his first wife, the long pain of Quine's divorce recorded in his autobiography is not mentioned here. Carnap hears indirectly about Quine's second marriage and sends him belated warm wishes and the unelaborated reassurance that "from my own experience I am a great believer in second marriages." There is a brief note of sympathy from Quine after Ina's tragic suicide much later in 1964, but nothing to indicate that Carnap ever said or wrote anything to Quine about it. This is primarily a philosophical correspondence.

Much of it consists of detailed responses of one author to the work of the other, or to the other's reactions to earlier responses. When Carnap is still in Europe they even exchange copies of grammatical and stylistic corrections made to letters written in the recipient's but not the author's native language; all are duly recorded here. Philosophically, the two go through each other's work line by line. Marginalia on the letters are also reproduced here, and a subsequent letter usually makes the point at greater length. In a small but reassuring number of cases, the objection is quickly conceded: a mistake was made. To follow the more complex responses for full scholarly understanding of the history of these matters the reader would also need at hand a copy of the text under discussion. Published writings of Carnap and Quine would not always suffice. Aside from differences of pagination, they are typically the outcome of these exchanges, not in every respect the object of them.

There is heavy concentration throughout on what look like questions of terminology. It starts in 1934 with Carnap seeking advice about English equivalents of crucial logical terms for the translation of *Die logische Syntax der Sprache.* By 1939 he is sending long questionnaires not only to Quine but to Alonzo Church, Ernest Nagel, Carl Hempel, Helmer, and Alfred Tarski, trying to get general agreement on a terminology for semantics and on English terms for a translation of his *Abriss der Logistik.* In 1943 he sends Quine and others another long questionnaire on the best terminology for various expressions and what he calls their "meanings," e.g., propositions, properties, concepts, individual concepts, and the like. This raises the suspicion that in some cases perhaps more than terminology has been at stake all along. Quine attempts to identify substantive disagreements between them on just these notions, but Carnap in all reasonableness does not "quite understand" what he has in mind; they look to him like different "proposals" for constructing a semantic theory. Very close study of these frank and careful exchanges would be needed to locate and identify the real sources of philosophical difference between Carnap and Quine. There is more direct give-and-take than in their finished published writings, but concessions made here and there for the sake of argument might also at times obscure differences that turn out to be important.

The heart of the book as it bears on this question is the rich one hundred pages which begin with a letter early in 1943, after a gap of two years with only a single postcard to show for it. Quine writes from Washington, where he is commuting daily from the outskirts to full-time work in Navy radio intelligence while apparently philosophizing by night. It is the time of the publication of his "Notes on Existence and Necessity." He reports that reflection on Carnap's recent *Introduction to Semantics* has led him to greater clarity on the points "crucial to semantics" at which he dissents from it. An earlier draft had been the subject of discussion in 1940 during that momentous period when Carnap and Tarski were both visiting Harvard (along with Russell, whose William James Lectures of that year became *An Inquiry into Meaning and Truth*). Carnap replies from El Paso and then from Santa Fe where he has been flat on his ailing back for many months while working continuously on what was to become *Meaning and Necessity.*

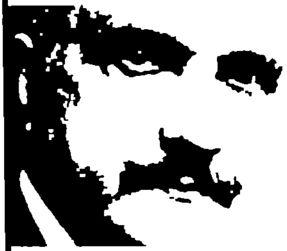
He first provides a "comparison of terminology" with expressions in the 'Quine' column paired off with their corresponding expressions in the 'Carnap' column, and after suitable translation finds their views on semantics to be "much closer to each other" than it had seemed. Quine replies with a masterly two-part letter occupying twenty of these pages in which he explains and defends his objections in "Notes on Existence and Necessity" and responds to Carnap's explanation of their differences about "meaning" and "designation." In the meantime Carnap has found what he thinks is a way to avoid all the puzzles Quine has raised about his approach to semantics by restricting himself now to "the concept of extension and intension" alone. When the new two-hundred-and-fifty-page typescript of that material reaches him, Quine reports that he "found it very interesting. By this I mean that I sat down with it on coming home from work and didn't let go of it, except for a brief meal, until after proper bedtime. By that time I had read two-thirds of it: it had a grip on me like a mystery thriller. This evening I finished it." He confesses to needing more time for the fundamental issues to "ferment" in him, but in what remains of that wartime evening he is still able to jot down eight more pages of "critical remarks." The subsequent letters back and forth lead up to a statement of Quine's objections which Carnap asks him to supply for inclusion in the published version of *Meaning and Necessity*.

What is finally most impressive about this correspondence, beyond the high quality of thought, is the energy and concentration sustained throughout, and the desire and above all the will on the part of both philosophers to get things right. In reply to a publisher's inquiry in 1961 about the advisability of translating and republishing Carnap's *Der logische Aufbau der Welt*, Quine, who greatly admires the book, cannot resist adding that "at important points it is simply wrong: objectively so, quite apart from questions of philosophical party." Philosophy is conducted in these letters in that same objective spirit. That is not something it would ever have occurred to either of these correspondents to point out to the other, or to request of him. But it must have been a continuing source of the great but unspoken respect they so clearly felt for each other.

BARRY STROUD

University of California/Berkeley

SANTAYANA, PRAGMATISM, AND THE SPIRITUAL LIFE



BY HENRY SAMUEL LEVINSON

"This work will cause many to reconsider the traditional rendering of Santayana and of much of the pragmatic movement. . . . In terms of highlighting the distinctive contributions of Santayana to American pragmatism, Levinson's account is a singular contribution to scholarship."—Herman Saatkamp, Jr.

"This splendid book is not only a major reinterpretation of Santayana, it is also a powerful challenge to our assumptions about American pragmatism. . . . It is surely one of the most important books on a single figure to have been written by a specialist in religious studies for a long time."—Jeffrey Stout

384 pp., $39.95

available at bookstores or from
THE
UNIVERSITY
OF NORTH
CAROLINA PRESS
Post Office Box 2288
Chapel Hill, NC 27515-2288

Toll-free orders: 1-800-848-6224

New from SUNY Press

THE TITLE OF THE LETTER
A Reading of Lacan

Jean-Luc Nancy and Philippe Lacoue-Labarthe
Translated by David Pettigrew and François Raffoul

"And, for a rigorous reading of Lacan I refer to the fundamental and indispensable book by Jean-Luc Nancy and Philippe Lacoue-Labarthe, *The Title of the Letter.*" — Jacques Derrida, *Le Facteur de la Vérité*

"I advise you to read a book called *The Title of the Letter*. It is with the greatest satisfaction that I have read it. I could not encourage its distribution enough. I can say, in a way, if it is a question of reading, that I have never been read so well." — Jacques Lacan

This book is a close reading of Jacques Lacan's seminal essay, "The Agency of the Letter in the Unconscious or Reason Since Freud," selected for the particular light it casts on Lacan's complex relation to linguistics, psychoanalysis, and philosophy.

181 pages • $10.95 paperback • ISBN 0-7914-0962-7

SUNY Press • c/o CUP Services • PO Box 6525 • Ithaca, NY 14851
Domestic & Canada add $3 shipping; *Foreign,* add $4 shipping
NY State residents add your sales tax • Canadian residents add GST
1-800-666-2211 (orders only) • VISA, MasterCard, American Express

The Collected Writings of Rousseau

Discourse on the Sciences and Arts (First Discourse) and Polemics

ROGER D. MASTERS and CHRISTOPHER KELLY, editors

The only uniform English edition of Rousseau's major writings continues with his first political work.

Volume 2 of *The Collected Writings* contains the entire *First Discourse*, contemporary attacks on it, and Rousseau's replies to his critics; it concludes with his summary of the debate in his preface to *Narcissus*. Many of these texts have not been available in English

"Kelly and Masters are among the leading Rousseau scholars in the world today, and it comes as no surprise to find that the second volume of their masterful English edition of Rousseau's works puts us once again deeply in their debt . . . Taken together with the valuable editors' introduction, this book affords English readers a genuinely revelatory and hitherto unavailable path to the heart of Rousseau's thought"
—Thomas Pangle, University of Toronto

Dartmouth College $40.00 cloth

Also available — Volume 1

Rousseau, Judge of Jean-Jacques: Dialogues. $40 00 cloth

The Reality of Meaning and the Meaning of 'Reality'

EDDY M. ZEMACH

A bold, original interpretation of Wittgenstein reverses a generation of relativistic readings to offer a fresh and realistic theory of Mind and Reality.

Traditionally, philosophers held that expressions that correspond to a meaning (a mental entity) are meaningful. Wittgenstein is most often read by contemporary philosophers to reject both theses. Eddy M. Zemach shows the error of a relativistic interpretation, especially when Wittgenstein's later work on the philosophy of psychology is fully considered. Combining his interpretation of Wittgenstein as a mentalist with the use of Substance Logic, and the idea that meanings and their referents are tokens of the same type, Zemach restores the traditional view.

The Reality of Meaning presents a clear survey of the major issues in contemporary philosophy of language and of mind. A lively, good-humored philosophical argument that is informative, topical, and well seasoned with common sense, it is a book for all readers of philosophy, psychology, and cognitive science.

Brown University Press $45.00 cloth

New England

University Press of New England • Hanover, NH 03755-2159 • 800-421-1561

The Logic of Reflection

German Philosophy in the Twentieth Century

Julian Roberts

Concentrating on the work of Frege, Wittgenstein, Husserl, the Erlangen School, and Habermas, Julian Roberts examines the debate between "analytical" and "continental" philosophers and suggests that the most urgent problems of both schools focus on the key issue of reflection.

"An enormously impressive piece of work."—William Outhwaite $30.00

Justice between Age Groups and Generations

Philosophy, Politics, and Society, Sixth Series

Edited by Peter Laslett and James S. Fishkin

In this book well-known political philosophers reflect on issues of justice between age groups and generations.

"This new entrant in a distinguished series is fully worthy of its predecessors. It will carry the debate on a difficult topic to a new plane of sophistication."—William Galston $30.00

Democracy and Deliberation

New Directions for Democratic Reform

James S. Fishkin

This innovative book proposes a new kind of democracy for the modern era. James S. Fishkin discusses the benefits of a "deliberative opinion poll," and he places it within the history of democratic theory and practice.

"A timely book that confronts a problem of major importance for democracies and that proposes an imaginative solution."—Robert A. Dahl $17.95

De Summa Rerum

Metaphysical Papers, 1675-1676

G. W. Leibniz
Translated with an introduction and notes by G. H. R. Parkinson

This book—the first in a new series that presents original texts by Leibniz with authoritative English translations on the facing pages—makes many of Leibniz's early papers available in English for the first time.

The Yale Leibniz, Daniel Garber and Robert C. Sleigh, Jr., series editors $35.00

Institutional Realism

Social and Political Constraints on Rational Actors

Robert Grafstein

Combining philosophy with an approach to social action drawn from rational choice theory, Robert Grafstein explores the nature of institutions and explains how these human products can influence—and sometimes coerce or constrain—their creators.

"An enormously useful and wide-ranging contribution to a debate that is of growing importance."—Ian Shapiro $30.00

Now available in paperback

Mental Imagery

On the Limits of Cognitive Science

Mark Rollins

The first philosophical study of mental imagery.

"An excellent introduction to the central elements of the imagery debate....A thoughtful discussion of the philosophical questions."—Michael J. Tarr, *Science* $12.00

Yale University Press
92A Yale Station, New Haven, CT 06520

Philosophical

Mystic Union

An Essay in the Phenomenology of Mysticism

NELSON PIKE

"Pike's book is the best treatment of Christian mysticism since Farges, Poulain, Butler, and Maritain in the early part of the century. It is a significant contribution to the literature on mysticism, and to the philosophy of religion in general." —William J. Wainwright, University of Wisconsin, Milwaukee. *Cornell Studies in the Philosophy of Religion.* $29.95

Radical Realism

Direct Knowing in Science and Philosophy

EDWARD POLS

In this eloquent and original book, Pols challenges the linguistic consensus that has dominated Anglo-American philosophy in this century. $28.50

Self/Power/Other

Political Theory and Dialogical Ethics

ROMAND COLES

Coles here explores the writings of Augustine, Foucault, and Merleau-Ponty in order to fashion an ethos that emphasizes the value of dialogical relationships between the self and others. $29.95

Knowing Words

Wisdom and Cunning in the Classical Traditions of China and Greece

LISA RAPHALS

Raphals here examines the role and significance of metic intelligence in classical Chinese philosophy, literature, history, and military strategy. $39.95

On Aristotle's *Prior Analytics 1.1–7*

ALEXANDER OF APHRODISIAS

Translated by Jonathan Barnes, Susanne Bobzien, Kevin Flannery, and Katerina Ierodiakonou. A leading Peripatetic philosopher of the early third century, A.D., Alexander of Aphrodisias provided a commentary on Aristotle's *Prior Analytics* which remains one of the most thorough and helpful guides to this work. *Ancient Commentators on Aristotle.* $47.95

Simplicius

Corollaries on Place and Time

TRANSLATED BY J. O. URMSON

The sixth-century Neoplatonist Simplicius here surveys ideas about place and time in the preceding thousand years of Greek philosophy and discusses late Neoplatonist theories, which he sees as representing a substantial advance on previous views. *Ancient Commentators on Aristotle.* $47.95

Primary Ousia

An Essay on Aristotle's Metaphysics *Z and H*

MICHAEL J. LOUX

Loux here presents a fresh reading of two of the most important books of the *Metaphysics*, Books Z and H, in which Aristotle presents his mature theory of primary substances (*ousiai*). $39.95

Inquiries

The Public and the Private in Aristotle's Political Philosophy

Judith A. Swanson

"One of the most accurate and comprehensive books on the *Politics* now available in English."—Thomas L. Pangle, University of Toronto. $32.95

Willful Liberalism

Voluntarism and Individuality in Political Theory and Practice

Richard E. Flathman

"In a set of fascinating readings, Flathman articulates the relevance for political philosophy of thinkers like Wittgenstein, William James, Nietzsche, Michael Oakeshott, William of Ockham and Duns Scotus."—Tracy B. Strong, Editor of *Political Theory*. $31.50 cloth, $13.95 paper

Dignity and Practical Reason in Kant's Moral Theory

Thomas E. Hill, Jr.

Hill addresses the most central concepts in Kant's *Groundwork of the Metaphysics of Morals*, as well as more specific moral issues drawn from his later ethical writings. $36.50 cloth, $13.95 paper

Matter, Space, and Motion

Theories in Antiquity and Their Sequel

Richard Sorabji

New in Paper! "As always, Sorabji displays much learning, and an excellent knowledge of ancient and modern physical theories. . . . A valuable contribution to problems in ancient philosophy and the history of science."—*Religious Studies Review*. $16.95

The Constitution of Rights

Human Dignity and American Values

Edited by Michael J. Meyer and William A. Parent

In this collection of original papers, twelve distinguished philosophers and legal thinkers focus on the relationship between our constitutional rights and human dignity. $32.50 cloth, $12.95 paper

In Their Best Interest?

The Case against Equal Rights for Children

Laura M. Purdy

"An important contribution to the debate about children's rights. Purdy's treatment provides an essential component of any adequate discussion of the topic; it is a useful corrective for the prevalent concentration on justice and rights."—Virginia Held, Department of Philosophy, Graduate Center, City University of New York. $32.50 cloth, $12.95 paper

At bookstores, or call 607-277-2211 (credit card orders only, please)

CORNELL UNIVERSITY PRESS, 124 Roberts Place, Ithaca NY 14850

Princeton

Moral Differences

Truth, Justice and Conscience in a World of Conflict

Richard W. Miller

In a wide-ranging inquiry Richard W. Miller provides new resources for coping with the most troubling types of moral conflict: disagreements in moral conviction, conflicting interests, and the tension between conscience and desires. Drawing on most fields in philosophy and the social sciences, including his previous work in the philosophy of science, he presents an account of our access to moral truth, and, within this framework, develops a theory of justice and an assessment of the role of morality in rational choice.

Paper: $16.95 ISBN 0-691-02092-2
Cloth: $65.00 ISBN 0-691-07409-7
Due July 1992

Hegel and the Spirit

Philosophy as Pneumatology

Alan M. Olson

Hegel and the Spirit explores the meaning of Hegel's grand philosophical category, the category of *Geist*, by way of what Alan Olson terms a pneumatological thesis. Hegel's philosophy of spirit, according to Olson, is a speculative pneumatology that completes what Adolf von Harnack once called the "orphan doctrine" in Christian theology—the doctrine of the Holy Spirit. Thus *Hegel and the Spirit* demonstrates the truth of Karl Barth's observation that Hegel is the potential Aquinas of Protestantism.

Cloth: $24.95 ISBN 0-691-07411-9

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540
ORDERS: 800-777-4726 • OR FROM YOUR LOCAL BOOKSTORE

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 8
AUGUST 1992

page
387 *Metaphysical Pluralism* Huw Price
410 *Rigidity, Ontology, and Semantic Structure* Alan Sidelle
431 *A Popular Presumption Refuted* V. H. Dudman

BOOK REVIEWS

433 Malcolm Budd: *Wittgenstein's Philosophy of Psychology* Colin McGinn
436 Michael Ayers: *Locke: Volume I, Epistemology; Volume II, Ontology* R. S. Woolhouse

441 NEW BOOKS: PAPERBACKS
441 NEW BOOKS: TRANSLATIONS
443 NEW BOOKS: ANTHOLOGIES

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice-President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1992 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER. Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO 8, AUGUST 1992

METAPHYSICAL PLURALISM*

David Hume is the St. Francis of modern metaphysics, the patron saint of ontological ascetics. The devout Humean takes the pinnacle of metaphysical virtue to be a world in which the only facts are the mundane first-order physical facts about how things actually are—"a vast mosaic of local matters of particular fact, just one little thing and then another," as David Lewis[1] puts it. Hume's own ontological austerity derives in part from his empiricism, in part from the great medieval Franciscan metaphysician, William of Ockham, and in part no doubt from the canny inclinations of his native land. Whatever its original sources, however, the doctrine remains enormously influential. The Humean ideal continues to exert great force in contemporary philosophy.

Like prosperous Franciscans, however, many metaphysicians who pay lip service to the virtues of ontological economy have strayed from the true path. The Hume world has frequently been judged too cramped. Some lapsed Humeans extend the bare Humean structure to fit in modal facts; some extend it to fit in subjective experience or intentional mental states; others to fit in moral facts; and so on. The common theme is that a respectable metaphysics simply cannot survive on the bare regime that Hume prescribes.

Now, the galling thing about a prosperous Franciscan is not the fact that he is not suffering the discomforts of poverty, or not this as such. It is his moral inconsistency, the fact that he claims virtue while failing to live according to his own professed conception of the virtuous life. There are two ways to challenge this inconsistency.

---

* I am grateful to audiences in Canberra, Brisbane, Sydney, and Wollongong, and particularly to John Burgess, John Campbell, John Hawthorne, Michaelis Michael, Philip Pettit, Michael Schepanski, and Mark Walker, for many comments on earlier versions.

[1] *Philosophical Papers: Volume II* (New York: Oxford, 1986), p. ix.

0022-362X/92/8908/387–409

One is to demonstrate to the Franciscan that he could make do with less, and hence establish that by his own lights he is not leading the pious life. This is in effect the strategy of those who dispute the need for particular extensions to the Hume world, arguing that the purpose such extensions serve can be met in some less extravagant manner.[2] The alternative deflationary strategy is more subtle. It is to accept the case for the extensions, for the Franciscan's additional consumption, but to point out that the consequence of accepting this is to undermine the minimalist conception of a virtuous lifestyle. In accepting that the Franciscan is fully entitled to satisfaction of the normal human appetites, in other words, we undercut his claim to abnormal piety.

It may not be obvious that there is any analogous move in the metaphysical case. In this paper I want to show that there is, and to suggest that this has profound consequences for our understanding of a range of contemporary metaphysical debates. Roughly, its effect is to undercut the distinction between various non-Humean forms of metaphysical realism and something akin to a Wittgensteinian linguistic pluralism. To the extent the distinction can be drawn, moreover, the latter is the default position. So, not only is this form of pluralism an important and widely neglected option in a range of contemporary metaphysical debates; it actually has claim to be the pre-eminent option. In a sense I shall explain, it is the philosophical geodesic, the course from which no one is entitled to depart without good reason. The paper thus presents a challenge to the lapsed Humeans of contemporary metaphysics: embrace worldly pluralism, or return to the pure faith, for there is no virtuous middle way!

The paper is in five main sections. Section I identifies our target species of pluralism, and distinguishes it from a more common species. In section II, I consider this target species in relation to a variety of antipluralist rivals, including in particular the monist doctrine that seems to be the intended position of the lapsed Humeans; and I draw attention to the central issue of the paper, namely, as to whether there is actually a tenable distinction between this monist position and our target pluralism. Section III argues that despite the avowedly ontological nature of their concerns, these lapsed Humeans cannot avoid reliance on a certain semantic distinction—essentially, the distinction between descriptive and nondescriptive uses of language—and points out that this puts them at at least a prima facie disadvantage compared to the pluralists, who need no such distinction. Section IV then outlines a case for thinking that the required distinction cannot be drawn, and draws attention to some connections between the resulting sort of pluralism and certain other recent approaches to the same metaphysical topics. And in section V, I illustrate the advantages and character of such a pluralism with reference to the issue between realists and their instrumentalist critics; perhaps surprisingly, this rather Wittgensteinian pluralism turns out to provide an especially secure brand of realism.

---

[2] A recent example is Bas van Fraassen's critique of contemporary non-Humean accounts of modality in *Laws and Symmetry* (New York: Oxford, 1990).

I. TWO KINDS OF PHILOSOPHICAL PLURALISM

One familiar kind of philosophical pluralism is exemplified by W. V. Quine's brand of ontological relativity, and perhaps in a different way by other forms of scientific relativism. Here the plurality consists in the possible existence of a range of alternative scientific worldviews, each empirically adequate to more or less the same degree, and none, even in principle, having privileged claim to provide a "truer" description of the world. This form of pluralism itself comes in several varieties or subspecies, of course. The distinction between Quine's variety and Kuhnian scientific relativism may well be a deep one, for example.[3] For present purposes, however, what matters is something that these views have in common, namely, that the plurality they admit involves a range of different ways of doing the same kind of thing, *of performing the same linguistic task*. There may be many equally valid possible scientific worldviews, but all of them are scientific worldviews, and in that sense are on the same level of linguistic activity. In other words, this is what might appropriately be called ***horizontal pluralism***.

Horizontal pluralism is far from confined to this scientific level. It is a product of certain kinds of relativism, and so may be found wherever these flourish. In ethics, for example, it is the familiar thesis that there is a range of equally coherent moral viewpoints, none objectively superior to any other. Why is this a case of horizontal pluralism? Again, because the plurality it envisages is confined to a single linguistic plane or level. Different moral systems are all nevertheless ***moral*** systems. They have something in common in virtue of which they may be counted to be different ways of performing the same linguistic task. It may be a very nice question how this something in common is to be properly characterized, but it must have an answer, if relativism is not to degenerate into the trivial point that the same words may mean different things for different people.

---

[3] On this see G. D. Romanos, *Quine and Analytic Philosophy* (Cambridge: MIT, 1983), ch 3.

If these are cases of horizontal pluralism, what would be a vertical pluralism? It would be the view that philosophy should recognize an irreducible plurality of *kinds* of discourse—the moral as well as the scientific, for example. This is the species of pluralism with which we are going to be most concerned. (I shall mainly use the term *discourse pluralism.*) I want to show that it provides a natural if not initially a congenial home for the lapsed Humean.

The plurality of forms of discourse, or "language games," is a prominent theme in the later Wittgenstein: "We remain unconscious of the prodigious diversity of all the everyday language-games because the clothing of our language makes everything alike."[4] There also seems to be a strong element of discourse pluralism in the American pragmatist tradition, of which Nelson Goodman and Richard Rorty are the most prominent recent representatives. True, Goodman and Rorty's pluralism is not exclusively vertical. It also contains a strong horizontal or relativistic element. The fact that the vertical component is significant in its own right is manifest, however, in the contrast between the positions of Goodman or Rorty, on the one hand, and Quine, on the other. Quine at times seems close to a purely horizontal pluralism, to the view that all factual discourse is either eliminable or reducible to physical discourse. This view acknowledges that there might be alternative physical discourses—"alternative physical theories, insusceptible to adjudication," as Quine[5] puts it—but says that there is nothing else at any other level. And certainly Quine rejects Goodman's proposal to admit on an equal footing a multiplicity of further "world versions," such as those of art and music, saying that in his view "this sequence of worlds or versions founders in absurdity" (*ibid.*, pp. 97–8).

More accurately, however, Quine himself is nonphysicalist to the extent of accepting the existence of certain abstract objects, such as classes and numbers. This is not to say that he intends to be a discourse pluralist. On the contrary, he is critical of the proposal to regard the acceptance of such objects as in some sense radically unlike the acceptance of physical objects:

> There are philosophers who stoutly maintain that 'true' said of logical or mathematical laws and 'true' said of weather predictions or suspects' confessions are two uses of an ambiguous term 'true'. There are philosophers who stoutly maintain that 'exists' said of numbers, classes and the like and 'exists' said of material objects are two uses of an ambigu-

---

[4] *Philosophical Investigations*, 3rd English ed. (Cambridge: Blackwell, 1968), p. 224; also §23–4.

[5] Quine, *Theories and Things* (Cambridge: Harvard, 1981), p. 98

> ous term 'exists'. What mainly baffles me is the stoutness of their maintenance. What can they possibly count as evidence? Why not view 'true' as unambiguous but very general, and recognize the difference between true logical laws and true confessions as a difference merely between logical laws and confessions? And correspondingly for existence?[6]

Later I want to argue that this point backfires. Discourse pluralism is the default position in this debate, and does not need to be defended by appealing to a claimed ambiguity in 'true' or 'exists'. On the contrary, it is the discourse pluralist's opponent who needs to appeal to some suitably "thick" or substantial unity in these notions —and Quine's preferred thin notions are not up to the task. Hence I want to suggest that Quine himself might best be recast as a discourse pluralist about abstract objects. He thus exemplifies what I shall argue to be the predicament of many would-be non-Humean realists in contemporary philosophy. (The striking thing about Quine's case is that his predicament stems from his own minimalism about truth and existence.)

The first task is to mark off discourse pluralism from a range of other ways of dealing with the same philosophical topics.

II. DISCOURSE PLURALISM AND ITS RIVALS

As a distinctive philosophical doctrine, discourse pluralism is perhaps best understood in terms of what it denies, in terms of the contrasts it tries to establish with other philosophical treatments of problematic topics. In this section, I describe four such contrasts. I distinguish a pluralist treatment of a philosophical topic from four other approaches: reductionism, two forms of irrealism, and lastly a view I call additive monism, which is the intended position of non-Humean metaphysicians. (Later our main interest will be in the question whether an additive monist can resist reconstitution as a discourse pluralist.)

To establish the general framework, consider a problematic topic against the background of an unproblematic topic—morality against the background of natural science, say. Assume for simplicity that the status of the background topic is not at issue. Our first group of distinctions rest on different combinations of answers to two questions about the status of the problematic topic. First question: Is its (purported) subject matter distinct from that of the background discourse? Second question: Is the problematic topic fully legitimate, again by whatever standards may properly be applied to the background discourse? (What is meant by 'legitimate' here will

---

[6] *Word and Object* (Cambridge: MIT, 1960), §27.

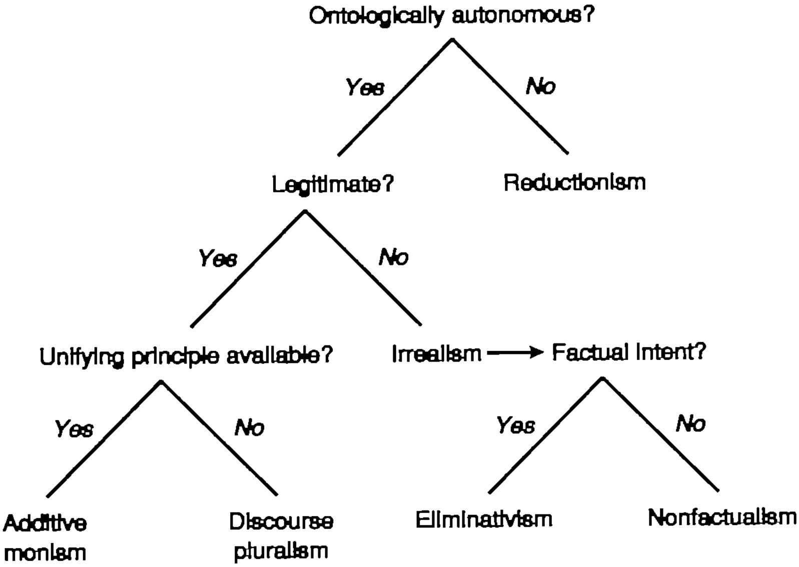


Figure 1: A discourse status quiz.

become clearer as we proceed.) In terms of our example then: Is the subject matter of moral discourse distinct from that of natural science?[7] And is it a fully legitimate discourse, or does it in some way fall short?

Discourse pluralism answers "yes" to both questions, and hence may be distinguished from views which answer "no" to one or other (and also from the less interesting view, which I shall not mention further, which answers "no" to both). The first contrast is therefore with *reductionism*, which concedes autonomy in order to save legitimacy. (We are at the first level in the tree shown in figure 1.) The reductionist agrees with the discourse pluralist about the legitimacy of moral discourse, but seeks to secure it by denying the ontological distinctness of its subject matter. On the contrary, the reductionist argues, morality is a branch of natural science—a fact that is obscured, in practice, simply by our usual ignorance of the appropriate identities between moral facts and natural facts. For the reductionist moral discourse is legitimate because it actually is natural discourse, albeit disguised natural discourse.[8] A pluralist wants both autonomy and legitimacy, whereas a reductionist is prepared to sacrifice the former in the interests of the latter. (The reductionist may not view it as a sacrifice, of course.)

---

[7] There has been much discussion as to whether there are autonomous levels of explanation in science. Those who argue that there are such levels commonly insist that, nevertheless, physics has ontological priority. Functionalists about the mental will often profess to ontological physicalism, for example, though claiming that psychology involves autonomous modes of description of (certain special arrangements of) physical entities or properties. For present purposes, I shall assume that this ontological criterion is capable of bearing the weight, and hence that such a view does count as a form of reductionism in the sense of figure 1 (in contrast say to Cartesian dualism). If this assumption were to prove untenable, then I think the defenders of multiple levels of description, like those I am calling lapsed Humeans, would be hard pressed to distinguish their view from a discourse pluralism about the scientific hierarchy.

Our next contrast is with two views that make the alternative sacrifice, conceding ground on legitimacy in order to preserve autonomy. They concede ground in very different ways, and we shall need to treat them separately. The fact that they both do so in some sense is, however, a reason to deal with them under a common heading.

Irrealism thus comes in two main forms. The first is *eliminativism,* the approach famously exemplified in J. L. Mackie's[9] *error theory* of moral discourse. In effect, the eliminativist holds that a problematic area of discourse *tries* to be legitimate, as well as autonomous; but that it fails to connect with anything in reality, and so is systematically false. The problematic topic is thus legitimate in intent, but fails in the execution of that intent. Mackie notwithstanding, eliminativism has not been particularly popular among moral philosophers. There are more familiar cases elsewhere, however. It has recently become a prominent approach to intentional psychology, for example. And Quine has long been seen as an eliminativist about meaning (and hence at least in some respects about psychology).

The second form of irrealism—*nonfactualism,* as I shall call it—is much more popular. It is the approach that preserves the autonomy of disputed topics by giving ground on their factual character. So, whereas the eliminativist holds that moral discourse tries to be factual but fails to connect, the nonfactualist denies that moral claims are ever intended to be factual claims. They do not fail to connect, for their linguistic role is not to attempt to connect. What it is instead is a matter of some interest, of course. In the moral case, the traditional options are the *expressive* or *emotivist* view, namely, that moral judgments express certain sorts of evaluative psychological attitudes; and the *prescriptivist* view, which treats a moral assertion as something approaching a command.

---

[8] Bear in mind that the relevant notion of reduction is an ontological one; see the previous fn.

[9] *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977).

Nonfactualism is a very common doctrine. Apart from the moral case, the approach is also well-known in application to aesthetic judgments, mathematical statements, theoretical sentences in science, secondary qualities, knowledge claims, psychology more generally, meaning ascriptions, indicative and subjunctive conditionals, and probabilistic, causal, and modal judgments. (No doubt I have missed some.) Not always well-appreciated are some of the basic philosophical commitments of any such position. It is clear that such a view is committed to the existence of a significant distinction between factual and nonfactual uses of language. Perhaps not so obvious is the obligation that the nonfactualist incurs as a result of this, namely, to account for the fact that nonfactual uses of language can present themselves as factual uses, at least superficially.

The contemporary philosopher who has done most to acknowledge and to try to meet this obligation is Simon Blackburn.[10] Drawing on Hume's idea that we project our attitudes and prejudices onto the world, and so see it as populated by seeming facts of our own construction, Blackburn has argued that the nonfactualist can explain our conversing as if there really were such facts. He has tried to show that the projectivist is entitled to a notion of truth, and to the other trappings of a realist linguistic practice. Thus, projectivism supports *quasi realism,* as Blackburn calls it. Couched in these terms, the nonfactualist's concessions concerning the legitimacy of a disputed topic are now very muted. To be sure, moral discourse (or whatever) is not *really* factual; but it has and is entitled to all the trappings of factuality, all the appearances of legitimacy. So, although on this view there is only one real world, only one realm of *genuine* facts, our language quite properly works as if there were many.[11] If this is not already discourse pluralism, then what separates it is the availability of a substantial distinction between factual and nonfactual uses of language. If that goes, nonfactualism is no longer a distinct alternative to discourse pluralism.

So much for the distinction between discourse pluralism and the various views that differ from it in denying either the autonomy or the legitimacy of a problematic area of discourse. We now turn to a contrast not marked by a disagreement on either of these issues, namely, the distinction between discourse pluralism and the doctrine I call *additive monism*. Additive monism agrees with discourse pluralism in rejecting reductionism and both forms of irrealism, accepting that multiple domains of discourse may each be autonomous and yet fully legitimate. The disagreement is only about how these separate domains are to be construed. The additive monist regards them as subdomains of a *single* universe of facts—not a single *physical* universe, presumably, for that would be to concede the game to a physicalist reductionism, but single in some sense. What distinguishes additive monism from discourse pluralism is the claim that there is something that unifies the various autonomous discourses; what distinguishes it from reductive physicalism, the claim that whatever this unifying principle is, it is not that all facts are ultimately physical facts.

---

[10] See in particular his *Spreading the Word* (New York: Oxford, 1984)

[11] Why not say "as if this realm were more inclusive than it actually is," so that the alternative to nonfactualism becomes additive monism? Here I anticipate a little—we shall see that the availability of a substantial distinction between factual and nonfactual uses of language turns out to be a requirement for the additive monist, as much as for the nonfactualist.

It is natural to wonder whether there a real distinction between discourse pluralism and additive monism, or whether the apparent issue rests on an empty metaphor. And if there is not a real distinction, how is the single resulting position best described? These neglected questions are of enormous importance to a wide range of current philosophical debates. For additive monism seems the intended home of the lapsed Humeans with whom we began—of contemporary non-Humean approaches to such topics as diverse as causation, modality, the mental, abstract objects, morality, and so on. On the face of it, these accounts argue that there are more kinds of facts in the world than the Hume world admits—facts about causal relations, over and above facts about constant conjunction, for example. The proponents of these views do not think of themselves as Wittgensteinian pluralists, of course. But it is not clear that there is conceptual space for any other position. The additive monist needs to show that something unifies the various autonomous discourses. Failing this, pluralism is the default option—what the would-be monist must fall back to if the distinction cannot be maintained. Thus, I want to argue that it is much harder not to be a Wittgensteinian pluralist than many people have assumed. Since a pluralism of this kind explicitly abandons Humean metaphysical minimalism, its effect here is deflationary—it undermines the Humean conception of metaphysical virtue. The issue turns on the availability of a unifying principle.

What might the additive monist appeal to at this point?

III. A PRINCIPLED MONISM?

A first thought might be that supervenience will do the trick. Perhaps additive monism corresponds to the view that a problematic topic is supervenient on a nonproblematic background (though not reducible to it); and discourse pluralism to the more extreme version of autonomy involved in denying both reducibility and supervenience between discourses. A little reflection shows that this will not do the job, however. There are counterexamples in both directions. Non-Humean metaphysicians commonly deny that facts about laws, causation, and the like are supervenient on first-order matters of particular fact.[12] Conversely, it is far from clear that a pluralist need deny supervenience. A nice counterexample may be extracted from Blackburn's discussion of moral supervenience.[13] Blackburn argues that his projectivist quasi realism not only accounts for moral supervenience, but is especially well-adapted to doing so. It is true that quasi realism is not a form of discourse pluralism. But I think it differs mainly in assuming that some stronger form of realism is appropriate elsewhere, for (at least some) nonmoral topics. So we would have a form of discourse pluralism if we abandoned this assumption, and thus extended quasi realism "all the way down." At the same time, we would still be entitled to Blackburn's account of moral supervenience, for this rests simply on the idea that moral discourse projects evaluative attitudes. Hence it seems that supervenience is independent of the issue between discourse pluralism and additive monism.

With supervenience out of the way, the next suggestion may be that additive monism is primarily a metaphysical doctrine, whereas discourse pluralism is primarily linguistic. The monist maintains that the autonomous domains of facts are all part of a single universe or *metaphysical* totality; whereas the pluralist seems to be denying the unity of factual *language*, opting instead for the view that a variety of different uses of language share certain superficial characteristics —Wittgenstein's common clothing of the diversity of language games. (What are these characteristics? More on this in a moment.) But although this certainly distinguishes the two doctrines in terms of their natural characterizations, it does not exclude the possibility that they come to the same thing. After all, the pluralist may well take the view that metaphysics itself is a manifestation of the superficial clothing of language, and derives its apparent unity from this very source. Thus, if metaphysics is essentially a concern with ontology, with "what there is," and this issue is seen in Quinean terms, then metaphysics and grammar become inseparable. Ontology is a matter of quantification, and this is precisely the sort of linguistic feature that the pluralist might take to be an element of the superficial clothing of language, serving similar grammatical purposes in discourses whose underlying functions in language are widely varied. (Analogous remarks would apply to truth.)

---

[12] See David Armstrong, *What is a Law of Nature* (New York: Cambridge, 1983), for example. It is my contention that these views would be better explicitly recast in terms of discourse pluralism, but clearly they are not intended as such.

[13] See "Moral Realism," in *Morality and Moral Reasoning*, Casey, ed. (New York: Methuen, 1971), pp. 101–24; *Spreading the Word;* and "Supervenience Revisited," in *Exercises in Analysis: Essays by Students of Casimir Lewy*, Ian Hacking, ed. (New York: Cambridge, 1985), pp. 47–67.

Monists might object at this point that they have a more substantial ontological principle in mind, and that their dispute with discourse pluralism turns largely on the question of our entitlement to some such notion. I think that this is on the right lines, but that it leaves out an important part of the story. The dispute cannot be entirely ontological, but must have a semantic or linguistic dimension. For the metaphysical view requires that there be a nontrivial sense in which the various discourses to which it applies are all serving the same linguistic function. It requires that they are all *descriptive* in the same sense, that they share a common goal of truth, a common concern to portray the facts as they are. It is not enough for the monist that there be a unified world out there; it is also crucial that within each disputed part of language, statements stand in the same relation to the relevant part of the single world. Otherwise, monism is trivial: it is easy to find a unified world to which every use of language relates *in some sense*. The monist requires that it always be *the same sense*. So monism needs a common semantic theory, as well as a unified world.

The monist might simply begin at this semantic level, of course, professing a concern with *truth, the facts,* or some such thing (again claiming to construe this semantic notion in a substantial way, incompatible with the pluralist's suggestion that these concepts belong to the superficial common layer of language). My point, however, is that even if the commitment does not begin at this semantic level, *it must get there in the end*. In effect, the monist needs a distinction between factual discourse—the kind that does or is intended to describe a part of a metaphysical totality—and nonfactual discourse, the kind that is not so intended.[14]

This obligation may seem to fall as much on the discourse pluralist as on the additive monist. I think there is an important asymmetry, however—the onus here lies predominantly with the monist. For

---

[14] Strictly, a monist might consistently hold that all discourse is factual (a position adopted, e.g., by Lewis in "General Semantics," in *Philosophical Papers: Volume I* (New York: Oxford, 1983), pp. 198–232). What matters is not the factual/nonfactual distinction as such, but the characterization of the factual itself

diversity is obvious, being guaranteed by difference of subject matter (once reductionism is rejected). It is the monist's *unity* that calls for substantial argument. So, not only does additive monism turn out to involve a linguistic doctrine after all, its monism requiring or entailing a linguistic characterization; but it also turns out to be a significantly more committed linguistic doctrine than discourse pluralism. The monist must hold that the unity of factual discourse is more than skin deep (or clothing deep, in Wittgenstein's metaphor). Whereas the pluralist can be content with an easy going linguistic "multifunctionalism"—with the view that under their common syntactic skin the roles of the various discourses in our lives may be as different as say those of our various internal organs.[15]

Additive monists thus incur a significant debt in the coin of semantic theory: they owe us a substantial characterization of factual discourse. (We have already noted that the same is true of nonfactualism.) I want to exploit this point to argue for the attractions of discourse pluralism over additive monism, or for that matter over its other nonreductionist rivals. In effect, I want to suggest that the debt thus incurred turns out to be impossible to repay in any of the acceptable currencies. We could accept payment in the debt's own currency, accepting the fact/nonfact distinction as an uncashed primitive. Ordinary scientific economy recommends instead that we give preference to a view that needs no such primitive. Pluralism and perhaps a physicalist reductionism emerge as the economical alternatives.[16]

This line of argument depends on a claim that many readers may find surprising, namely, that there is not a readily available and well-founded distinction between factual and nonfactual uses of language. I have defended this claim at length elsewhere,[17] and in the next section I sketch some of the main points of that argument, and some related advantages of discourse pluralism. I think the claim should really not be surprising to a contemporary philosophical audience, however, for in a sense it is a natural consequence of Quinean skepticism about meaning. Factuality is a semantic property, if anything. It is the property of "statementhood," or that property which is shared by all *assertoric* utterances. Thus, it is a particular component of meaning, and Quine's general skepticism about meaning might already have led us to recognize the possibility that there is no such property, in any substantial sense.[18]

---

[15] These also exhibit striking superficial similarities, looking to the untrained eye like so much raw meat. Who could have guessed that the kidney was really a filter, the brain a thinker, etc ? The pluralist is suggesting that the functions of language might exhibit a similar diversity and modularity.

[16] Physicalism might after all emerge as a degenerate form of discourse pluralism, namely, as the view that in fact there is only one autonomous and legitimate discourse, though there could have been many.

[17] In *Facts and the Function of Truth* (Cambridge: Blackwell, 1988).

IV. TRUTH AND FACTUALITY

I have said that discourse pluralism and additive monism are distinguished by the fact that the former makes no appeal to—indeed, explicitly rejects—the assumption that there is a more than superficial category of fact-stating or descriptive uses of language. It may seem perverse to count this a point in the pluralist's favor. For is it not obvious in what such a category consists? Fact-stating utterances are just those which have genuine truth conditions, those which have assertoric force, those which express genuine beliefs about that world. Surely these notions are unproblematic?

The pluralist's response to this challenge is to draw attention to the possibility that these common-sense intuitions might rely on nothing more than the superficial features of language, the clothing that covers a more fundamental diversity. Are the notions of truth, fact, assertion, belief, and so on foundational categories, inevitably central to any theoretical account of our use of language? Or are they mere products of language, categories thrown up by language itself, and not therefore presupposed by a proper explanatory theory of language? As a first step, the pluralist asks us simply to acknowledge that the latter answer is at least a conceptual possibility —that common sense is powerless to exclude it. Thus, an appeal to common sense is ruled out by the consideration that common usage is the very object of theoretical inquiry here. All sides are agreed that ordinary usage exhibits a superficial unity between the discourses in question. The issue is whether this superficial unity is more than skin deep.

Perhaps surprisingly, it is the pluralist who has the better of the naturalistic high ground in this dispute. The issue concerns the structure of a complete *naturalistic* theory of language use. Roughly, it is the question whether such a naturalistic theory reveals a structural feature of language that may be regarded as constituting a fact-stating/non-fact-stating distinction. This does not mean that the pluralist must concede that *all* linguistic notions are ultimately naturalistic (which would be to embrace a form of reductionism that the pluralist is in general keen to avoid). The pluralist simply wants us to recognize the possibility that what can be said about language in naturalistic terms might reveal no trace of a substantial factual/ nonfactual distinction. As a naturalistic question, this not the sort of thing that can be decided a priori. So discourse pluralism is ultimately an empirical doctrine, albeit a highly theoretical one, and one that concerns the linguistic part of the natural world.

---

[18] The odd thing is that Quine himself never seems to notice that his skepticism has this consequence. As we noted at the end of section II, he takes his minimalist intuitions about truth to support the view that there is a single category of descriptive discourse, *contra* those who would take an instrumentalist view of talk of abstract objects. But he fails to notice that this is the sort of "victory" which we achieve by sweeping the pieces from the board. The default position is a kind of pluralism—not the nonfactualist's kind, but the Wittgensteinian doctrine that superficial unity of our talk of truth masks wide differences in the functional roles of discourses or "forms of life."

The onus lies with the additive monist to show not only that there is a substantial property of factuality underlying ordinary usage, but also that it has a particular distribution. The monist wants to show that the disputed discourses are unified by a common concern, such as to "depict the world." The discourse pluralist is not denying that scientific utterances and moral utterances are alike in being meaningful speech acts, but simply that there is any more substantial sense in which they are both "statements of fact," or "descriptions of how things are." So a plausible strategy for the pluralist is to try to argue that none of the usual attempts to characterize such a property is capable of settling the issue of distribution; or in particular of excluding the trivializing possibility that all discourse is factual discourse—i.e., that "statementhood" is a property guaranteed to be possessed by any utterance whatsoever. If a proposed analysis of "statementhood" cannot exclude even this possibility, it is unlikely to convince us that the monist has a defensible claim concerning the factual status of the discourses of interest.

This skeptical attempt to trivialize any proposed analysis of the notion of "statementhood" is the strategy I adopted in *Facts and the Function of Truth*. To illustrate, the most common suggestion is that factual discourse is distinguished by its relation to the notion of truth. Genuinely factual utterances are said to be those which are genuinely truth-bearing, or genuinely "aimed at truth," or some such. As several writers have pointed out, however, the notion of truth involved had better be capable of doing the work. One notion which obviously will not do the work is a redundancy or disquotational notion, which depends on nothing more than the standard equivalence principle: the fact that for any sentence *P*, '*P*' and '*P* is true' seem in some strong sense to say the same thing.[19] So, if truth is to be invoked at this point, we need what Bernard Williams[20] called a *substantial* theory of truth. I argue in *Facts and the Function of Truth* that there is no such theory to be had. The contenders inevitably fail on one of two counts: either they are like the disquotational theory, in applying equally well to any sort of discourse; or they already assume a fact-stating/non-fact-stating distinction, in some other form.

---

[19] As we have already noted, Quine's minimalist views about truth fall into this category.

[20] "Consistency and Realism," in *Problems of the Self* (New York: Cambridge, 1973), pp 187–206, at p. 202

One such alternative relies on the psychological distinction between beliefs and other kinds of propositional attitudes. This proposal originates in the kind of emotivism which grounds itself on Humean moral psychology, drawing a contrast between descriptive or belief-expressing language, on the one hand, and expressive or evaluative attitude-expressing language, on the other. The emotivist treatment of moral discourse has provided a model for parallel treatments of other topics. In the process it has given rise, at least implicitly, to a general theory of the nature of the distinction between factual and nonfactual language: to the view that factual language expresses belief, whereas nonfactual language expresses attitudes of other kinds. The trouble with this theory is that there is very little prospect of drawing the required psychological distinction, in advance of some other solution to the problem it is supposed to address.[21] We need a substantial notion of belief, but we will not get that until we have a substantial way of drawing the same distinction somewhere else.

The skeptical attempt to drive a conventional doctrine around a logical circle has a famous paradigm in modern philosophy, in Quine's own attack on the analytic/synthetic distinction. And as we have noted, the conclusion itself seems one that Quine might well have endorsed. Quine's irrealism about meaning seems very much at home with the view that there is no substantial unity to the factual, descriptive, or truth-bearing part of language as a whole; no single such semantic category, in any substantial sense. Indeed, this view might be seen as simply doing for semantics what Quine himself does for ontology: insisting that philosophy has no privileged second-order vantage point, but must rather make do with the deliverances of the best first-order theories, taken at face value.

At any rate, the effect of these skeptical arguments is to underline the economical advantages of discourse pluralism, by showing that a useful characterization of factual discourse is not as readily to hand as we have tended to assume. But it might be objected that pluralism incurs costs of its own. For one thing, it owes us an account of what it regards as the superficial unity of language—for example, of the fact that all discourses apparently avail themselves of the same indicative syntax, and of the notion of truth. For another thing, we are entitled to be told in what the underlying *diversity* is supposed to consist—in what dimension the different discourses enjoy their vertical separations. Related to the latter point is the important question as to how we set the limits to pluralistic tolerance. Clearly there are some conceptual conflicts that call for a less tolerant response, for simply discarding one of the alternatives. The obvious thought is that this is appropriate when the theories concerned are attempting to occupy the same linguistic role, or level. But what does this mean, in general?

---

[21] In the original emotivist case, this difficulty has recently been highlighted by the inconclusiveness of several attempts to exclude the possibility that desires or evaluative attitudes might themselves be represented as a special class of beliefs. See, in particular, Michael Smith, "The Humean Theory of Motivation," *Mind*, XCVI (1987): 36–61, Philip Pettit, "Humeans, Anti-Humeans and Motivation," *Mind*, XCVII (1988): 530–3; Lewis, "Desire as Belief," *Mind*, XCVII (1988): 323–32; John Collins, "Belief, Desire and Revision," *Mind*, XCVII (1988) 333–42; Michael Smith, "On Humeans, Anti-Humeans and Motivation: A Reply to Pettit," *Mind*, XCVII (1988). 589–95; my "Defending Desire-as-Belief," *Mind*, XCVIII (1989): 119–27, and Philip Pettit and my "Bare Functional Desire," *Analysis*, XLIX (1989): 162–9.

These are real obligations for a discourse pluralist. I think they should not be unwelcome, however. On the contrary, they provide a further opportunity for pluralism to display its comparative creditworthiness, by showing that it has the means to meet such conceptual debts. In *Facts and the Function of Truth,* I tried to provide a suitable account of truth for these purposes—an account that has the potential to explain why diverse uses of language should be alike in being treated as "truth-bearing." The basic idea is that despite the diversity of uses of language, different uses tend to be alike in this respect: within each use or discourse, there is some potential utility in noting and resolving disagreements between the members of a speech community. A notion of truth encourages a community to realize this potential. It does so by generating a social value that is negative when speakers disagree and positive when they agree. Disagreement thus becomes socially unstable, and argument, with its long-run benefits, is thereby encouraged.

This approach does not require that argument is always beneficial, of course. Like any evolutionary theory, it keeps its eyes on the long run. Nor, importantly for a pluralist, does it require that the kind of benefit always be the same. The pluralist can thus allow, on the one hand, that there is no uniform way in which each of the many language games affects our well-being; and yet, on the other, that in all or most such games there is *some* benefit in argument, and hence in the availability of a notion of truth. In fact, I think the pluralist can do considerably more than this. For it turns out that our use of the notion of truth is *not* strictly uniform across the range of different domains of discourse. There are significant differences, underlying the predominant pattern. I think that pluralism can do much to account for these variations, by appealing to the different functions or linguistic roles of the different domains concerned. And *only* pluralism has the flexibility to do this: other approaches are tied too rigidly to the existence of a fact-stating/non-fact-stating dichotomy in language.

What are these discourse to discourse variations? They turn on the differential tendency of arguments to be resolved in such a way that neither participant is judged to have been initially mistaken—for example, the kind of thing that may happen when a disagreement about a probability turns out to have stemmed from the fact that the speakers concerned had access to different evidence. This kind of outcome to an argument manifests itself as a sudden reluctance to apply the notions of truth and falsity to the judgments concerned. Such no-fault disagreements occur for different reasons in different parts of language. To a surprising extent, however, it seems to be possible to explain where they do occur in terms of an understanding of the distinctive functional role of the type of discourse concerned, coupled with the general principle that the origins of truth and falsity lie in their role in encouraging useful argument. (In the process, incidentally, it is possible to explain many of the intuitions that have motivated nonfactualism—but to do so without the nonfactualist's problematic division of language into fact-stating and non-fact-stating categories.)

This approach has affinities with recent work of Crispin Wright's[22] on what might be called the "fine structure" of truth. Wright has distinguished a number of different components that may, but need not all, be characteristic of the use of truth in association with a particular area of discourse. In effect, he suggests that we may classify a subject matter according to which of the set of these characteristics we take its notion of truth to involve. As I understand it, his concern is mainly descriptive and taxonomic. I think there is some prospect, however, that the structure thus discerned will turn out to be explicable as sketched above, in terms of some general account of the function of truth in language. Whether the best general account will be in the terms I outlined earlier remains to be seen. But if it is to be in keeping with Wright's program, I think it will have to be like my account in being *explanatory* rather than *analytic:* its focus will be not on the question "What *is* truth?" but on the question "Why do ordinary speakers have such a notion as truth?"

Interestingly, Wright too takes this stance on truth to support a noncommittal metaphysics, at least as the natural fallback position.

---

[22] See "Realism—the Contemporary Debate: Whither Now?" forthcoming in *Reality, Representation and Projection*, J. Haldane and C. Wright, eds. (New York: Oxford).

He characterizes this position as an antirealist one, in Michael Dummett's sense, on the grounds that the minimal notion of truth it requires can be thought of as derived from assertibility. I disagree at this point, arguing that in practice the minimal notion of truth is already realist in this sense, for reasons to do with the origins of negation.[23] That issue aside, I take the following remark of Wright's to be in much the same vein as my emphasis on the economical advantages of discourse pluralism: "Anti-realism thus becomes the natural, initial position in any debate. It is the position from which we have to be shown that we ought to move. All the onus, everywhere, is on the realist" (*op. cit.*, p. 11, typescript). Substitute 'pluralist' for 'antirealist' and 'nonpluralist' for 'realist', and these are my sentiments exactly.

The project also has affinities with Blackburn's quasi realism. The quasi realist about moral discourse (for example) wants to argue that, although moral judgments are not *really* factual, they are entitled to the trappings of factuality, including a respectable notion of truth. So, although there is only one domain of *genuine* facts, it is quite proper on this view for ordinary usage (and truth in particular) to work as if there were many. As we noted earlier, what separates this from discourse pluralism is the quasi realist's assumption that his project is a limited one, bounded in its application by the availability of a substantial distinction between factual and nonfactual uses of language. If that distinction lapses, quasi realism is no longer a distinct alternative to discourse pluralism. Discourse pluralism is again the default position. But so long as a more limited quasi realism remains a live project, the quasi realist's interest in explaining why certain discourses should usefully employ an "artificial" notion of truth will apparently coincide with the project described above.

In sum, there are considerable grounds for doubt as to whether a useful principle of semantic unity is available to an additive monist, other than as an ad hoc and costly theoretical primitive. And there are considerable grounds for optimism concerning the project to explain what superficial unity we find in language (and the limits of this unity) on foundations that only the discourse pluralist finds congenial. The case is not closed, of course, but it is the additive monist who is on the defensive.

V. PLURALISM AS DEFENSIBLE REALISM

At this point it may be helpful to emphasize that discourse pluralism is not an irrealist position. The pluralist accepts with all sincerity that there are moral states of affairs, possible worlds, numbers, or whatever. What she rejects is the additive monist's attempt to put a further metaphysical gloss on such existential claims. (That gloss turns out to depend on a semantic distinction of questionable standing, or so the pluralist argues.) Without the gloss, discourse pluralism sits quite happily with a nonmetaphysical or "minimal" realism.[24]

---

[23] See my "Sense, Assertion, Dummett and Denial," *Mind*, XCII (1983): 174–88; and "Why 'Not'?" *Mind*, XCIX (1990): 221–38.

Perhaps more surprisingly, this species of realism turns out to be particularly well-protected against one of realism's traditional opponents. Discourse pluralism has a defense against an instrumentalist challenge, of a kind that continues to be pressed against realists in a number of branches of contemporary philosophy. This defense is simply not available to the monist, for, as we shall see, the antidote the pluralist is able to employ against instrumentalism works equally well against additive monism.

Instrumentalists prey on the Humean tendencies of the modern philosophical public, offering a simple path to ontological asceticism. To avoid the realist's ontological excesses, they tell us, we need simply treat our *talk* of such ontology as a convenient instrument. The talk works just as well when construed this way, and yet in ontological terms it is absolutely free. Thus, vast ontological savings compared to realism, and at no extra cost elsewhere.

Of course, realists often respond that the offer is too good to be true. There is a cost elsewhere, albeit one that may not be obvious at first sight. What we lose under instrumentalism, the realist claims, is the ability to explain why the instrument works. The need for such an explanation thus becomes a debt that the instrumentalist cannot pay off.

This response is a good deal less effective in some cases than in others, however. Whatever its merits as an objection to instrumentalism about theoretical physics, it is likely to be of little use in defense of realism about possible worlds, for example. Here the realist will typically acknowledge that a proper explanation of why talk of nonactual worlds is useful can only appeal to the actual world—i.e., in effect, acknowledge that such talk would be just as useful if there were no other worlds. Similar considerations might apply in the case of intentional psychology. Realists might well grant that an adequate explanation of why talk of beliefs and desires is useful could in principle be cast in terms of scientific theories making no reference to intentional states.

In cases such as this the realist needs a different strategy, and might be tempted to appeal to Quine. For has Quine not taught us that there is no more to the question whether there are possible worlds (say) than the matter as to whether such things are quantified over in the most serviceable philosophical theory of modality and related topics? And given that the instrumentalist does not deny us our *talk* of possible worlds (quantification included), do we not therefore have Quine's authority for rejecting the instrumentalist's claim that there are no such things? Does not the talk itself provide all the warrant we ever have for an existential claim?

---

[24] This point was made vivid to me in discussions with John Campbell. Discourse pluralism seems compatible with what Campbell calls the "simple view" about entities or properties of a given kind. See, e.g., his "A Simple View of Colour," in *Reality, Representation and Projection.*

It is important, however, to appreciate that there are two ways to interpret this argument, resting on two quite different ways of interpreting Quine's views on ontology. On one reading, the Quinean doctrine is effectively a principle of ontological quietism—the principle that there is no separate second-order science of ontology, but simply the mundane business of existential quantification carried out by first-order specialists in the course of their working lives. I shall come back to this: I think it does provide an objection to instrumentalism, but not the objection the metaphysical realist was hoping for. There seems to be a tendency to read the Quinean doctrine in quite a different way, however. Under this second reading, the activities of first-order specialists provide raw data to which the Quinean principle is then applied, yielding second-order ontological conclusions. In other words, the fact that first-order specialists quantify over the entities of a certain kind is held to constitute *evidence* that such entities actually exist. Far from dismissing the science of ontology, Quine's doctrine thus becomes the main instrument in the working ontologist's tool kit—a kind of "magic eye" for detecting otherwise invisible existents.

Had Quine provided us with such a magic eye it would indeed provide a powerful weapon against instrumentalists (or at least those instrumentalists not prepared to take issue with Quine). "See for yourself," the realist would say, inviting the instrumentalist to peer at reality through Quine's lens. But I think that to make this explicit is to make it obvious that Quine intends no such thing. The right reading of Quine is the quietist one.[25] Now this reading too counts against the instrumentalist. The instrumentalist requires that not all good theories involve ontological commitment (via their existential quantification); for it depends on whether the theory in question is genuinely descriptive or merely an instrument. As a quietist principle Quine's doctrine contends that there is no such further issue.

---

[25] This misinterpretation of Quine seems to me to parallel a misunderstanding concerning the role of best explanation of science. It is part of the practice of science to accept the best current explanations of observed phenomena. This is hardly more than a truism: to say that an explanation is the best we have is *ipso facto* to indicate that we give it more credence than anything else on offer. There is a tendency to make a second-order principle of inference out of this: to say that *because* they provide best explanations, the theories concerned are likely to be true, and therefore worthy of a further degree of credence. Another similar mistake is that of thinking that the principle that the best choice is the one that maximizes one's expected utility assigns a *further* value to the choice in question, over and above its expected utility.

But does this favor realism? Yes, but only a realism of the discourse pluralist's minimal kind. For what the quietist principle denies us is something that is needed by additive monists, as well as by instrumentalists. Instrumentalism is a form of nonfactualism. Like additive monism, though for a different reason, it thus depends on a substantial distinction between factual and nonfactual uses of language. The monist needs an account of what keeps the domains of discourse together; the nonfactualist an account of what sets them apart. The instrumentalist thus owes us an account of what is lacked by those domains of discourse which are judged "merely" instrumental, and hence of what is not lacked by those which pass the test. Without such an account, for one thing, there seems nothing to stop the above kind of argument for instrumentalism from going all the way down. All discourse would thus be construed as fictional discourse, and the contrast that gives point to instrumentalism would be lost. Modal and moral discourse would be on the same footing as discourse of any other kind, just as the pluralist and the additive monist both contend.

So the instrumentalist needs a substantial factual/nonfactual distinction. Often this is supposed to be cashed in ontological terms. The instrumentalist is presented as denying the existence of a class of entities of a certain kind—possible worlds, or mathematical objects, for example. We have just encountered a reason, however, for doubting that this can serve as the fundamental distinction. Construed in a Quinian sense, ontological commitment is a more or less trivial consequence of serious theoretical commitment; even if there is always room for argument as to which ontology strikes the best ideological bargain. Hence instrumentalism cannot be primarily an ontological doctrine. The instrumentalist needs a prior distinction between full-blooded and merely fictional theoretical commitment. An ontological distinction would flow from such a distinction, but cannot provide it; not, at least, unless we are prepared to discard the Quinean principle that our only guide to ontology is the quantificational structure of accepted theory.

All of this applies equally to the additive monist. Quine's ontological quietism prevents the monist from casting the unifying principle he requires in ontological terms.[26] There is nothing to prevent a discourse pluralist from embracing the Quinean doctrine, however; the pluralist, too, will be interested in the most advantageous theoretical formulation of any given level of discourse, and may well see the choice in Quinean terms. The pluralist thus has an objection to instrumentalism which would be suicidal in the hands of an additive monist.

---

[26] I argued in section III that the monist's unifying principle cannot be entirely ontological in any case, but must involve a semantic component. The present

Discourse pluralism thus provides a particularly defensible form of realism—a position uniquely placed to exploit some of the hidden costs of irrealism. A corollary is that those who find themselves in the grip of realist intuitions about matters that are otherwise vulnerable to instrumentalist attack would do well to take pluralism seriously, for it may well be the only realism they can have.

VI. CONCLUSION

We began with two ways to respond to the moral inconsistency of a prosperous Franciscan. One response was simply to point out that he could well make do with less. I compared this to challenging the lapsed Humeans' case for particular metaphysical luxuries. This is what nonfactualists do, in effect, in arguing that in order to make sense of say modal discourse, we do not need to assume a realm of modal facts; instead, we may read the discourse nondescriptively, so that its interpretation does not require such facts. My main goal in this paper has been to offer an alternative deflationary challenge to non-Humean metaphysics. This compares to granting that the Franciscan has the same range of human needs as anyone else, and the same right to reasonable satisfaction of those needs; but insisting that in granting this, we overturn the idea that there is any special virtue in self-imposed poverty.

Discourse pluralism thus allows the lapsed Humeans their talk of possible worlds, modal facts or whatever, subject only to the modified Quinean requirement that such an ontology provides the most economical basis for the particular discourse in question (i.e., roughly, the basis that best enables that discourse to serve its intended function in human life). What it withholds is the concession that such talk maps the bare objective structure of a single independent reality. On the contrary, we pluralists maintain, the idea that there is such a mapping rests on a theoretical error about language, namely, the view that there is a single substantial category of descriptive or fact-stating discourse.[27] It follows that the Humean conception of metaphysical virtue was always misguided. In granting the lapsed Humeans their metaphysical comforts, we thus deny them

---

point is that Quinean quietism undercuts the ontological strategy at first base, as it were.

[27] We pluralists think that many of those who try to deflate the lapsed Humean in the first way also fall prey to this misconception

their Humean metaphysical virtue; for we say that there has never been any such thing.

Pluralism thus takes the heat out of contemporary metaphysical debates about a very wide range of topics: modality, mathematics, moral discourse, conditionals, and many others. As a final example, a well-motivated pluralism might allow one to endorse something like Daniel Dennett's[28] account of the role and origins of intentional psychology, without having to concern oneself as to whether the view amounts to instrumentalism about the mental.

Some readers will see this supposed advantage as like that of theft over honest toil—or perhaps more accurately, of idleness over honest toil. I see it as rather the advantage of idleness over unnecessary and unproductive toil—the advantage of stopping work on a misconceived and unproductive philosophical seam. True, we cannot be certain that the seam will remain unproductive, but there is enough doubt about the matter for the wise course to be to consider some of the general issues on which its eventual productivity might depend. I have emphasized that these are naturalistic issues, in particular issues about language as a natural phenomenon. (Does it admit a significant descriptive/nondescriptive distinction, for example?) So they are hardly issues that the pluralist's likely opponents may justly ignore.

At the very least, we could then go back to work with a better idea of what we were looking for in the first place. Instead, however, we might want to explore other philosophical seams, other projects that would now seem more promising: perhaps the elaboration of the general descriptive/explanatory approach to truth; perhaps the broadly Kantian examination and classification of different domains of discourse, with reference to their role in human life;[29] perhaps the exploration of nice issues concerning the connections and disconnections between the various discourses. A different philosophy, certainly, but not an idle one.

HUW PRICE

University of Sydney

---

[28] See, e.g., "Real Patterns," this JOURNAL, LXXXVIII, 1 (January 1991): 27–51.

[29] Recent work attempting to generalize the notion of a secondary quality might be seen as falling under this heading. See for example Wright, "Realism-the Contemporary Debate: Whither Now?"; and Mark Johnston, "Objectivity Refigured," in *Reality, Representation and Projection;* Johnston's "Dispositional Theories of Value," *Proceedings of the Aristotelian Society, Supp. Vol. LXIII* (1989): 139–74; and Pettit, "Realism and Response-Dependence," *Mind*, C (1991): 587–626. For a critical view of this program, offering an alternative more in keeping with discourse pluralism, see my "Two Paths to Pragmatism," in *Response-Dependent Concepts*, Peter Menzies, ed. (Canberra: ANU Research School of Social Sciences, 1991), pp. 46–82

RIGIDITY, ONTOLOGY, AND SEMANTIC STRUCTURE*

The notion of rigid designation stands squarely at the interface of metaphysics and philosophy of language.[1] Itself a semantic notion, it is metaphysical in its applications, and in that whether an expression is rigid depends on whether its reference is in all worlds the same. Rigidity has played an important role in the development of a broad philosophical view that has largely supplanted the dominant empiricism of the earlier part of this century. On this new view, important questions about what is necessary, what features of things are essential, and what the conditions of identity and individuation are for many things are factual, indeed empirical questions. They require not conceptual or semantic analysis, but are rather to be settled by looking at the world, investigating the objects in question, rather than the expressions used to refer to them. Our findings tell us about the "deep" structure of the world, not about our conceptual scheme. Rigidity plays a central role in all this, because it is the fact that many key terms are rigid which allows us to fix upon an object so as to investigate it with respect to both the actual and other possible worlds. This new view, for which Kripke and Hilary Putnam[2] are largely responsible, is partly characterized by the following claims (these may not all be distinct views):

(1) By using expressions rigidly, we can ask counterfactual questions about objects without any a priori specifications concerning the properties of the object of reference.
(2) We can stipulate possible worlds without worrying about or settling questions concerning transworld identity conditions.
(3) Our modal inquiries are directed at the *objects*[3] *themselves,* and in no way concern our concepts or the meanings of words.
(4) Thus, we can make a posteriori *de re* modal discoveries—we can have empirical discovery of essence. These findings represent language and mind-independent modal features of objects.
(5) We can form identity statements between rigid designators

---

* Many people have helped me with the issues in this paper. Thanks to Thomas A. Blackson, David O. Brink, Berent Enc, Ted Everett, Martha Gibson, Paula Gottlieb, Elliott Sober, Dennis Stampe, and Leora Weitzman

[1] Saul Kripke introduces the notion in *Naming and Necessity* (Cambridge: Harvard, 1980), pp 48–9, and provides a fine overview in his preface, pp. 3–15.

[2] Putnam's views are presented in many of the essays in *Mind, Language and Reality* (New York: Cambridge, 1975); "The Meaning of 'Meaning' " is probably the most important.

[3] To ease exposition, I shall use 'object' to range over any object of reference or inquiry; it includes, e.g., kinds as well as individuals.

0022-362X/92/8908/410–30

(i) if the expressions happen to tag the same object, the statement is true, and necessarily so;
(ii) if we can only find out the truth of this statement empirically, we again have a posteriori necessary truths,
(iii) metaphysically and semantically, there is nothing more to the truth of an empirical identity statement than "same thing tagged."[4]

(1)–(3) contrast with the more traditional view according to which reference is determined through meaning. On that view, one cannot sensibly talk about an object in another possible world except as it satisfies, in that world, the conditions given by the meaning of some word. But then it is specified a priori that the referent of the word, whatever it is, has those properties, and to specify a possible situation is to specify one in which an object with those properties is in some situation, thus settling the issue of transworld identity. And since our possible world descriptions are constrained by the meanings of our words, we have no independent grasp of the object to investigate *it*—all we are really investigating are conceptual connections within our framework. (1)–(3) represent the more "world-" or "object-directed" understanding of these sorts of philosophical inquiry. (4) and (5) are two more overtly metaphysical features of the new view, exemplified by the purportedly necessary a posteriori 'Water is H_2O', 'Margaret Truman has Bess Truman for a biological parent', and 'Hesperus is Phosphorus'. The "object-directed" understanding of (1)–(3) feeds the metaphysical, rather than "conceptual," understanding of these phenomena, as represented in (4) and (5). I shall not discuss or criticize these positions individually; they are introduced here simply to point the reader toward a position that I hope is familiar, and understood to have a great deal of philosophical currency.

In this paper, I shall investigate how rigidity plays such an important role in contemporary metaphysics. I shall argue that the role depends on a certain understanding of rigidity, which I call the *metaphysical understanding* (sections I, II): this is the view that the semantics of rigid expressions is determined by mind-independent facts of transworld identity. I shall also offer a more semantical understanding of rigidity, with which this may be contrasted (section VI). The metaphysical understanding, I argue, requires a privileged ontology according to which some, but not all, ways of carving up the world get at "real" objects (sections III, IV). The semantic interpretation, by contrast, carries no such metaphysical burden. While I shall not argue against privileged ontologies, I think it important to see this commitment. Many who accept the new view may not realize that they are so committed, and may not be inclined to accept such an ontology. Further, anyone who wishes to buy into the currently prevalent picture which (1)–(5) partially characterize must defend this ontology. It is not at all supported, as one might think, by our intuitive judgments about which terms are rigid (section V). Indeed, I hope to show that it is by tacitly *assuming* such an ontology in familiar presentations of rigidity-supporting intuitions that we are so easily led to this larger metaphysical picture: the ontology, not our judgments of rigidity, does the work (section V). I leave it to defenders of the new view to try to provide the necessary arguments; I also hope my semantic account of rigidity will be of interest whatever one's ontological views.

---

[4] Of course, this is not the *content* of an identity statement on this view. It is just all that there is to such a sentence's being true

I. RIGIDITY *DE JURE* AND *DE FACTO*; REFERENCE THROUGH SENSE AND IDENTITY

An expression is rigid if it has the same reference with respect to all possible worlds (in which it refers). Kripke distinguishes two ways an expression may be rigid, *de jure* and *de facto* (*op. cit.*, p. 21 fn.). These correspond to two semantic interpretations for expressions, most easily illustrated with definite descriptions.

The statement 'The President of the U.S. is necessarily male' has two familiar readings: on one reading, it ascribes an essential property to George Bush; on the other, it states the impossibility of a nonmale attaining to a certain political office. Semantically, 'The President of the U.S.' functions differently in the two cases.[5] In the first, the description picks out some individual in virtue of its descriptive content, and is understood to refer, in other possible worlds, to *that individual* (Bush), and *its* counterfactual properties determine the truth value of the statement. The descriptive content of the expression falls away; it is relevant only in determining an *actual* referent. So used, we may say that such descriptions gain their referents in other worlds *through identity,* i.e., through the identity of objects in other worlds with the actual referent.

On the other reading, the description also picks out an individual by its descriptive content, but it picks out individuals in other worlds *by the very same content.* In each world, the individual that satisfies this descriptive condition (having a certain office) is its referent.

---

[5] While one can represent this by a scope distinction, there remains the question of what in the utterance of a modal sentence with descriptions makes it properly interpreted as having wide or narrow scope.

Descriptions so used may be said to gain their referents in other worlds *through sense,* i.e., through the meaning of the description.

A description that refers counterfactually through identity will obviously be rigid, since the very condition for the expression's applying to something in another world is that that thing be identical to the actual referent. The rigidity of such expressions is rigidity *de jure;* the function of such an expression with respect to other worlds is precisely to pick out that very object. Descriptions that refer counterfactually through sense, on the other hand, may, and typically will, turn out to be nonrigid, for the objects that satisfy the conditions in various worlds may not all be the same, as with 'The President of the U.S.' In some worlds, it is George Bush, in others, Michael Dukakis or Elizabeth Dole. Now, it can turn out that the referents *are* all the same, as with 'The sum of two plus two'. Such expressions will be rigid, but, so to speak, by accident.[6] Their rigidity is rigidity *de facto.* The semantic function of the term, as regards other worlds, is given by a descriptive condition—it is not *stipulated* that the referent in other worlds is to be the actual referent. It just turns out that way.

These distinctions are not confined to descriptions. For any referring expression useable in counterfactual discourse, we can ask how it functions semantically (through identity or sense), and can classify its rigidity, if any there be. Of course, if an expression *has* no descriptive content, no semantic structure, all that can determine its counterfactual reference will be the actual referent itself; such "contentless" expressions will consequently refer through identity and be rigid *de jure.* According to direct or causal theories of reference, names are like this, and they provide the paradigm of rigidity *de jure.*

Most philosophers have not taken much interest in rigid *de facto* expressions. The reason, I think, is that they cannot play the metaphysical role exemplified in claims (1)–(5). Since a rigid *de facto* expression has its counterfactual reference determined through sense, the answers to important modal questions will be analytic and a priori. When 'The President of the U.S.' is used this way, it is given a priori that the referent in any world will be the president there. This also constrains counterfactual stipulations we can make using such expressions—we cannot stipulate worlds in which the referent fails to have these sense-specified properties. And this constraint is given a priori (unlike our inability to stipulate worlds that violate necessary a posteriori truths). So here, our modal investigations are not so "object-directed" as when reference is through identity. When reference works through sense, even when rigid, our means of getting an object of modal inquiry already build in, in *making* it the object of inquiry, what counterfactual properties it must have.[7] The lack of interest in rigidity *de facto* is the (implicit) recognition that, for the metaphysics, rigid reference must occur through identity.

---

[6] One *can* intend such a description to refer counterfactually through identity, making its rigidity not "accidental," but *de jure.* With such 'identifying' descriptions, we may fail to notice that they can be rigid in both ways It is easier to see for 'the element with atomic number 79' (assuming that used through sense, this rigidly picks out gold).

And for metaphysics, then, the relevant distinction is not that between rigid and nonrigid expressions per se, but that between expressions which refer counterfactually through identity, and those which refer through sense.[8] Names have pride of place because (on the current view) they always refer through identity, while descriptions only sometimes do. Note, though, that names need not be *wholly* without content to be rigid *de jure*. All that matters is that this content plays no role in determining their reference *in other worlds*. If names need to, or sometimes do get actual referents by reference-fixing descriptions,[9] they will function just like descriptions used rigidly *de jure:* both will have their *actual* referents determined through sense, but their counterfactual referents determined through identity.[10]

---

[7] A "sense" based view can allow *some* modal questions to be left open—for counterfactual conditions may be specified generally or abstractly (e g , 'same origin' or 'same deep structure'), leaving empirical investigation to find out what *in particular* (actually) "fills in" these conditions. This means that a (modified) traditional view can allow necessary a posteriori truths, and "stipulation" of other worlds need not (implicitly) answer *all* modal questions. But what is left open would have *modal* import only because of the semantics, undermining the supposed *metaphysical* import of (1)–(5) mostly given by (3), the second sentence of (4), and (5) (iii). For examples of such "general" specification, and more on a "conventionalist" account of a posteriori necessity, see my *Necessity, Essence and Individuation* (Ithaca: Cornell, 1989), esp. chs. 2, 3.

[8] Or, alternatively, between expressions that are rigid *de jure* and all others. As the mechanism is what matters, I prefer my first formulation. The importance of this distinction is also noted by A. D. Smith, "Rigidity and Scope," *Mind*, XCII, 369 (April 1984) 177–93.

[9] See Kripke, *Naming and Necessity*, pp. 55–6 for his introduction of the reference-fixing/meaning-giving distinction.

[10] Thus, a causal account of reference is not, as such, vital to the "new" metaphysics. Further, while Kripke has answers to description theorists who claim that names are descriptions which happen to always take wide scope in modal contexts, their position would anyway not touch the metaphysical side of his work, as the senses of such descriptions fall out as irrelevant to determining counterfactual reference.

II. THE METAPHYSICAL UNDERSTANDING OF RIGIDITY

The familiar picture of how rigid designation allows for empirical inquiry into necessity and essence actually makes plain that the importance of rigidity resides in rigidity *de jure* and reference through identity. The story is that we introduce a term to refer rigidly to some object. This term then gives us a handle on that entity so we can investigate its properties, not only actually, but also, because of rigidity, in other possible worlds. The subject of our investigation is then *that object,* and any positive results we find give us modal properties of *it.*[11]

This picture is, I hope, familiar enough. But as we have seen, if the *object* is truly to be the subject of our inquiry, it cannot just be rigidity at work here. What is key is not simply being able to refer to some object in other possible worlds, but being able to do so in a way that *does not build in what features it will have* in these worlds. We need *the object itself,* and not some description, to determine the expression's counterfactual reference. Expressions that refer counterfactually through sense do not meet this condition, even when they are rigid (*de facto*).

Philosophers who *do* accept claims (1)–(5) then, accept, or are committed to, a particular understanding of how rigidity works, at least in the relevant cases (those in which we get empirical discovery of necessity). We can characterize this commitment first by a negative claim, *the descriptionlessness thesis:*

> DL: (Many)[12] rigid designators have no descriptive content determining (or constraining) their application in counterfactual situations.

Our earlier remarks make clear why we do not say simply 'have no descriptive content'. Descriptive content that merely determines *actual* reference, as in rigid *de jure* descriptions, is no bar to the sort of object-directed modal inquiry in which we are interested.

But how, then, *is* counterfactual application for such terms determined? It is through identity: the counterfactual referent just *is* the actual referent. I shall call this positive claim *the metaphysical understanding of rigid designation:*

---

[11] Such investigations might have turned up empty. So, as Nathan Salmon has argued (in "How *Not* to Derive Essentialism from the Theory of Reference," this JOURNAL, LXXVI, 12 (December 1979): 703–25; and *Reference and Essence* (Princeton: University Press, 1981)), the theory of reference, or the notion of rigidity does not *itself* give the metaphysical results. But it renders such investigations sensible, whereas on traditional theories, we can only investigate objects as they satisfy the conditions given in the sense.

[12] The qualifier is just to acknowledge rigid *de facto* descriptions.

> MRD: The reference of (many) [see fn. 12] rigid designators in other worlds is given metaphysically—it is whatever object in the world in question is (identical to) the actual object of reference.

I call this the "metaphysical" understanding because once one has determined an actual referent (and perhaps has intended that the expression be rigid), the world takes over—the facts about identity across worlds determine counterfactual reference.

Whether MRD is the same as the claim that many rigid designators are rigid *de jure*, or have their reference determined through identity, depends on how we choose to use these latter notions.[13] Whatever we decide about that, MRD is worth separate presentation, since it states explicitly what is required for rigidity to do its metaphysical work. MRD asserts not only that some expressions have their rigidity semantically built in, but further, that this happens *metaphysically;* it is simply the mind/language independent fact that x in w_n is identical to y in the actual world that makes x the referent of T in w_n. This takes a definite stand both on there being such independent facts, and on our ability to fix our terms' referents without this reference-fixing playing the crucial role in determining counterfactual reference (in some way other than on traditional theories); this is left open, at least on one reading, by the notions of rigidity *de jure* and (perhaps) reference through identity.

I hope that many readers recognize MRD as the view that they *do* have about how rigid designation works (except for rigid *de facto* descriptions); I have mostly just tried to articulate it, and to explain why those who accept the metaphysical view that usually goes with presentations of rigidity are committed to it.

My next task is to show that MRD—and so (1)–(5)—requires a "privileged" ontology, and that the phenomenon of rigidity does nothing to support such an ontology or, consequently, MRD.

---

[13] Since it is so typically assumed that "nonaccidentally" rigid terms always lack descriptive conditions for counterfactual reference, the issue of whether rigidity *de jure* is *defined* by "built-in" rigidity, or also requires nondescriptive reference, does not arise. As we shall see (sec VI), however, a term can have built in rigidity despite having descriptively determined counterfactual reference. So we shall have to decide. If rigidity *de jure* is just built in rigidity, then it does not suffice for MRD, and MRD is a *theory* about how such terms work. If we decide instead that descriptive reference precludes rigidity *de jure*, it is the same as MRD—though it is no longer "data" that names, say, are rigid *de jure*. Similar remarks apply for reference through identity.

The important question is: Do names, and expressions that function in modal contexts like names, have their counterfactual reference determined by mind-independent facts about the identity of objects across worlds? This is the question of whether MRD is true. Whether it is *also* the question of whether there is in fact rigidity *de jure*, or reference through identity, is but a terminological matter

III THE STRANGE AND UNUSUAL

Let me describe a purported entity, which I shall call "The President of the U.S.," or "Prez" for short. Prez is currently composed by George Bush, was composed in 1790 by George Washington, and in 1863 by Abraham Lincoln. In some other worlds, Prez is currently composed by Mike Dukakis, and in others by Elizabeth Dole. A person comes to constitute Prez, usually, by being sworn into office, though it may not work this way in all other worlds. Prez bears basically the same relation (at a time) to the people that are (or have been, or could be) Prez that a statue or river bears to the portions of matter (clay, water) that constitute them (at a time).[14] Prez is not a mereological sum of the people who have been (are) the president (unless these other entities [statues, rivers] are, which they are not —at least of those things which *actually* constitute them, supposing mereological sums to have their parts essentially). When someone leaves office, Prez does not go out of existence, but rather, comes to be constituted by another person, as the Nile comes to be constituted by different (portions of) water.[15]

Of course, Prez differs from statues and rivers in that it is not continuous through space and time—when there is a swearing in, Prez jumps in space to occupy the location of the person who has just been sworn in. Between such ceremonies, though, Prez follows a person-like path of spatiotemporal and psychological continuity. Fully precise identity conditions for Prez would be tricky, but I hope the basic idea is clear. Prez is constituted by, and not identical to, the people who hold the office of President of the U.S. And even in worlds where the presidency is abolished after Washington, Washington would *be* Prez only in the sense of constitution, as they would still differ modally, and in their identity conditions.

Prez is person-like in that it has mental states and performs intentional activities ("The President of the U.S." vetoes bills); however, as just noted, it does not have the identity conditions persons have (say, spatiotemporal or psychological continuity). For a general category, we may call Prez an "office person"; other office people are "The Owner of the Yankees" and "The Pope." Office people also belong to broader class of (purported) entities, which includes "My Dog," which was composed by Fido, and now by Spot, and "Elizabeth Taylor's Husband," which I leave to the reader to articulate.

---

[14] On constitution, see David Wiggins, *Sameness and Substance* (New York: Oxford, 1980); and "On Being in the Same Place at the Same Time," *Philosophical Review*, LXXVII, 1 (January 1968): 90–5.

[15] The Nile also comes to be constituted by different *water*, but here I am concerned with the relation between individuals.

For reasons which should be plain, I shall call all of these "unusual" entities; for now, it should suffice to say that unusual entities are ones that we do not usually (explicitly, at least) recognize as (single) entities.

Let us call an ontology that recognizes unusual entities a *permissive* ontology, and one that does not *restrictive*. Further, let us call an ontology that recognizes *no* objects—one according to which there is stuff in the world, but none of it endowed with the sort of identity conditions we think distinguish objects from mere stuff—a *fully restrictive* ontology. We may then call an ontology that is neither fully restrictive, nor permissive, a *privileged* ontology. In a privileged ontology, some but not all ways of carving up the world get at real objects; some but not all proposed identity conditions are the identity conditions for some class of objects.[16] It is this sort of ontology to which defenders of MRD are committed.[17]

As I said, I shall not argue against a privileged ontology, but it is incumbent upon the defender of such a view to say what the favored entities have, and unusual ones lack, in a way that neither leads us to wholesale restrictivism, nor just makes us want to introduce a new general term to include both favored and unusual entities. As a middle position, it needs to withstand challenges from both sides, and it is not obvious that acceptable answers to one of the other positions can be made which will not push one to the opposite view. While there are many differences between people and office people, it is very questionable whether any of these has the right sort of *metaphysical* significance.[18]

IV. WHY MRD REQUIRES A PRIVILEGED ONTOLOGY

(a) Restrictive ontology. It should be plain that one cannot advocate the metaphysical understanding of rigidity with a wholly restrictive ontology. Since on this view there *is* no (mind-independent) identity across possible worlds, the facts of identity just are not up to the

---

[16] Of course, everyone agrees that conditions like 'horse with a horn' pick out no real objects. Whether a 'way of carving' gets at real objects rather concerns its *individuative* force *given* its actual application. Wholesale restrictivists allow that there are "portions of the world" which are composed of H_2O, but deny that, mind-independently considered, this condition is individuative, or the essence of anything

[17] In fact, their ontology must rule out not only the "unusual" entities under discussion, but quite a few not-so-unusual entities, since *any* widespread spatiotemporal overlap of objects will raise the problems of indeterminacy to be discussed. Our "unusual" entities just make the problem more vivid.

[18] Eli Hirsch offers a thorough challenge to privileged ontologies in "Rules for a Good Language," this JOURNAL, LXXXV, 12 (January 1988): 694–717. A rare explicit defense of a privileged view, or at least, attack on extreme permissivism, occurs in Sydney Shoemaker, "On What There Are," *Philosophical Topics*, XVI, 1 (Spring 1988): 201–24

task of determining the counterfactual reference of a term—indeed, even *actual* reference is problematic. If one *tried* to introduce a term as rigid *de jure,* to refer in other worlds to whatever is identical with the actual referent, the term would just fail to refer in other worlds. And obviously, the metaphysics (1)–(5) characterize could never get off the ground. Put simply, MRD presupposes mind-independent crossworld identity, and with a wholly restrictive ontology, this supposition is false.

Despite this obvious incompatibility of a wholly restrictive ontology with the new metaphysics, it is worth noting. For if MRD also cannot abide a permissive ontology, then the defender of MRD cannot just defend a restrictive (as opposed to a permissive) ontology, but further, this defense cannot lead back to extreme restrictivism. The metaphysical understanding of rigidity requires a privileged, and not merely a restrictive, ontology.

(b) Permissive ontology. Can rigidity be metaphysically determined in a permissively populated world? I shall first consider the rigidity of descriptions, and then apply our findings to the case of names.

As standardly discussed, descriptions may be used either rigidly or nonrigidly. If 'The President of the U.S.' refers through sense, it will denote Bush nonrigidly; if it refers counterfactually through identity, it rigidly designates Bush. But now suppose we allow Prez in our ontology. Referring through sense will no longer suffice for 'The President of the U.S.' to be nonrigid, for in virtue of its descriptive content, it can pick out Prez just as well as Bush, and if it picks out Prez, it does so in every possible world—for Prez is in every world and always the President of the United States. Indeed, one might think that Prez, if it exists, is the more natural referent of 'The President of the U.S.', so on a permissive ontology, *every* definite description, if used to refer through sense, threatens to be rigid *de facto.* For now, though, all that matters is that on a permissive ontology, descriptions used to refer through sense, without further specification, are indeterminate between rigidly denoting unusual entities, and nonrigidly denoting the particulars that constitute them at a given time.[19]

On the other side of the coin, using our description to refer counterfactually through *identity* leaves it open whether the value of 'The President of the U.S.' in other worlds is Bush or Prez. For again, without further specification, they have equal claim to be the actual referent of the description, but, of course, they determine different referents in other worlds. To designate Bush rigidly with this description, it will not suffice simply to intend to refer counterfactually through identity.

---

[19] This is not to say such descriptions would always *be* indeterminate—only that when they are not, it is not just in virtue of intending to refer counterfactually through sense

In both cases, a permissive ontology threatens the determinacy of reference. In the case of referring through sense, this appears as a puzzle about how we can manage to use descriptions nonrigidly; in the case of referring through identity, the problem is determining which of (at least) two things is rigidly designated. Whereas on a wholly restrictive ontology, there are too *few* mind-independent crossworld identities (i.e., none) for MRD, on a permissive ontology, there are too *many*. MRD presupposes not only that there *are* some identities, but, since it asserts that counterfactual reference is determined simply by the fact that some object in another world is identical to the actual referent, that it is possible to refer determinately to some actual object in a way that allows this.

Now, it *is* possible to get determinate reference even within a permissive ontology. Since the objects the reference hovers between differ in their identity conditions, we can determine reference to one of them by intending to refer to the object with such-and-such identity conditions. Unfortunately for MRD, this is not a way to secure reference which just lets the "facts of identity" determine counterfactual reference—for here, this reference is determined *semantically*, not metaphysically: the counterfactual referent is whatever satisfies the intended conditions. By associating identity conditions with uses of expressions, we can also make out, despite the above problems, how there can be nonrigid uses of descriptions, even with a permissive ontology. This will be explained in section VI.

It should be clear that these same problems apply when we turn to names. On the standard view, we get a referent for a name—via ostension or description—and then let this referent guide counterfactual reference. Clearly, when a name is introduced by a description, our position is just as found above for the rigid use of descriptions—simply intending the facts of identity to determine counterfactual reference underdetermines such reference. But the same holds true for more purely ostensive introductions of names. Within a permissive ontology, wherever we point, there are too many overlapping entities to which we may be referring, entities which differ in their identity conditions and modal properties. Until we specify what is to count as the *same* thing, we have no determinate reference to any one thing at all. Of course, we *do* achieve determinate reference with, say, 'Bush'. But this leaves completely open *how* this occurs. The argument is not that 'Bush' cannot be determinate and rigid, but only that for it to be so, it must have more semantic content than just its referent. So again, there is no way to let the *thing*—the referent—determine the counterfactual reference of the name rather than, as in the case of descriptions, the criteria of sameness (identity). For a name to be rigid, it must be associated with a criterion of identity across worlds, and this then serves as the criterion of application for the name. While there can be rigidity, MRD again fails.

To sum up, MRD involves two presuppositions: first, that there are mind-independent facts of crossworld identity; and, second, that we can get determinate reference in the actual world in a way that allows these facts of identity to determine reference in other worlds. On a wholly restrictive ontology, the first (and maybe the second) is false; and on a permissive ontology, the second is false. Both suppositions can be satisfied only within a privileged ontology; consequently, MRD, and the philosophical picture that requires it, require such an ontology. Some who accept these views may be happy with this consequence—but the burden is on these "new" metaphysicians to defend this ontology. They cannot just say, "Well, since it is entailed by our position it is as well supported as the rest of our view." For the principle evidence for the view starts from our intuitions about rigidity, and as we have seen, it is not these data *themselves*, but their metaphysical interpretation that is the cornerstone of the new view. Since this interpretation presupposes a privileged ontology, one is not entitled to it, or what one thinks flows from it, without *first* defending the ontology. But one might think this defense need not be wholly independent. Perhaps our judgments of rigidity *themselves* support a privileged ontology. I shall argue in the next section that this is not so.

V DO OUR DIFFERENTIAL JUDGMENTS OF RIGIDITY SUPPORT A PRIVILEGED ONTOLOGY?

To get people to see the rigid/nonrigid distinction, we ask similar questions about different terms. If we look at counterfactuals using the name 'Nixon', it seems that all are used to talk about "this man"—the same one we speak of in actual situations. On the other hand, if we look at 'The President of the United States', it seems that, in other worlds, it may pick out someone else. Asked, of all the counterfactual referents of this expression, if they are the same, we naturally say "no." In saying no, we assert a nonidentity, and it is a short step from here to supposing that there is no such entity as "The President of the United States." At least, if we accepted a permissive ontology, would we not be likely to say that 'The President of the U.S.' *was* rigid, that its referents *were* identical? And if we were restrictivists, would we not *deny* that 'Nixon' was rigid, that *its* referents are identical?

Thus, our common judgments of rigidity and nonrigidity might be thought to lend support to a privileged ontology, or at least show that most people accept this ontology. And one might think this at least puts the burden of proof onto the view's opponents.

This argument from our differential judgments about rigidity to a privileged ontology is problematic. I do not think our judgments support the claim that we *do* accept such an ontology, must less the claim that we *should*. For one thing, if one were to *ask* people who make the familiar judgments about nonrigidity whether or not they thought, say, that "The President of the U.S." existed, I think many of them would at least admit that they had no special reason for denying this. "Sure, why not?" would be a common response, I am guessing. They need not be thinking of such entities when the questions of rigidity are raised; one may need to ask explicitly. This might lead them to reconsider their earlier judgments that 'The President of the U.S.' is not rigid, in which case the "phenomena" might vanish. But they need not. One may think that while "The President of the U.S." populates the universe, it is not the referent of many (or any) uses of 'The President of the U.S.' That 'The President of the U.S.' does not always have the same referent—or that we do not think it does—is immediately just a fact about what we are referring to—not what there is (or is not). In section VI, I provide an account of how this description could be used nonrigidly (and rigidly) to denote particular people, even if there were "The President." So judgments of nonrigidity, of themselves, can hardly show that we are committed to a privileged ontology. Further, and perhaps more importantly, I think we can give a plausible reason why people would make the nonrigidity judgments they do even if, on reflection, they were the firmest of permissive ontologists (or, suitably modified, the firmest of extreme restrictivists).

The questions we ask when attempting to decide whether a term is rigid or not are not in any obvious way questions about what entities exist. Generally, they are identity questions, e.g., 'Is the same entity present in all relevant situations'? Or 'Is the referent of the term in all worlds the same'? So it should be no surprise, first off, if these inquiries do not pique people's ontological sensibilities. This point is highlighted by other philosophically familiar uses of variables like 'thing', 'object', and 'entity'. Asked flat out, it is familiar that questions and statements involving such variables are deeply ambiguous.

'How many things are there on the table'? asked of a table containing a deck of cards, may legitimately receive a number of answers, most obviously 'one'—a deck of cards or 'fifty-two'—cards; or who-knows-how-many combined cards, numbers, spots, etc. These answers require no special permissiveness in one's ontology—we all recognize cards and decks as among the world's occupants. But still, to get a definite answer—or ask a definite question—we need something more specific than 'thing'. Interestingly, we use such general terms all the time, successfully and without problem. Particles of dust and cookie crumbs, for instance, are understood not to be counted when one asks 'How many things are there on the table'? This question ordinarily means, roughly, 'How many medium-sized dry goods are on the table'? My not counting cookie crumbs in no way shows that I doubt their existence or status as "things." When we use these most general terms, we have more or less agreed upon "default settings" that specify more particularly the type of thing in which we are interested. We may have different defaults for different contexts and, certainly, one can specify directly at any time. The important point is that we *need* not verbally explicitly do so. And because these pragmatic factors can be tacit, our ordinary intuitions about these things hardly show much about our ontology.

Given that we have such conventions, or conversational customs, concerning these most general terms, it is not hard to see how they might enter into our investigations concerning rigidity. Indeed, we need not bring in "The President of the U.S.," and the like, to see that they do. Consider our question of whether the same entity is present in any situation relevant to 'George Bush is necessarily male'. Nobody even *thinks* to say "Well, there is one collection of molecules the instantiation of maleness of which is *actually* relevant, but in other worlds, there are different such collections." There is no denial implied here of the ontological "thingness" of these collections—they are entities, all right—it is just supposed that we are talking about medium-sized dry goods like the man, George Bush.

Just so, I think it is with our nonrigid inquiries. Consider the reading of 'The President of the U.S. need not be male' which we defend by pointing to a world in which Elizabeth Dole is the President. When asked whether this use is rigid—Does the same entity figure in all the relevant situations?—we answer "No," since it is Bush who is relevant in the actual world, and Libby Dole in the world that proves the claim. Does this mean we accept, or would accept if asked, a privileged ontology in which "The President of the U.S." does not appear? No more, I think, than ignoring the cookie crumbs when counting the things on the table implies a similar rejection of cookie crumbs. In the context, we naturally interpret the question as 'Does the same *person* always appear'?—indeed, in practice, the discussion often takes place in these terms: 'The referent is not always the same *person*'. And to this, the answer is clear. This is also why collections of molecules do not figure in our intuitive reasoning. (The question of rigidity, however, hinges on sameness of *entity;* so this finding, however clear, does not even suffice to show that in this sentence the expression *is* used nonrigidly, much less that a restrictive ontology is indicated.) Thus, I am not at all convinced that the fact that people can be led so easily to recognize definite descriptions as nonrigid shows that, at any level, they accept a privileged ontology. It only shows that we are accustomed to speak of a certain subclass of entities—even to the exclusion of other entities that we clearly do recognize.

These reflections upon why we should not make too much ontological hay of our ordinary judgments of rigidity—or other judgments that may play on our "default" settings for general variables—also help explain why we are so naturally led to a metaphysical interpretation of rigidity and, ultimately, to the new metaphysics. Our ordinary practices—our "default" settings—have us speak, and perhaps believe, "as if" we accepted a privileged ontology. Unless these conventions are somehow flagged, we make our judgments *within* these practices, and may, without realizing it, incur more blatantly metaphysical commitments—in this case, to a privileged ontology.

Now, part of the great elegance and beauty of Kripke's work—especially his discussion of rigidity—is that we are explicitly asked *not* to "think as philosophers," but to offer our common-sense judgments about how we use names and descriptions. Thus, in getting us to see the rigid/nonrigid distinction, and that names, but not all descriptions, are rigid, no "metaphysical flags" are raised. One would have to be predisposed toward ontological axe-grinding to think *at that point* that the question 'Does "The President of the U.S." always refer to the same thing'? was a "loaded" question. We have not yet incurred a privileged ontology, but our attention is firmly fixed on "usual" entities (people), and not at all on "deeper" ontological questions. As we move on to the metaphysics, and wonder how rigid expressions work, we still avoid the relevant sort of metaphysical flags. Names, we are told, are introduced by ostension or description, and are then used rigidly for 'that thing'—whatever we have ostended or described; ditto for rigidly used descriptions. Now, if we are on our ontological guard, we may at this point complain that ostension or simple description will not suffice to pick out (determinately) a single, modally extended object. But there is no indication yet, nothing really to warn us, that we *should* be on such guard. Ostension is a perfectly normal practice; we know how, without needing to make a fuss, to identify 'that thing'. This *could* be because we accept a privileged ontology, but it could as easily be because we have default settings here which we use all the time—without, again, even noticing it. And by now, without having had to confront explicitly any metaphysical questions—just working uncritically within ordinary linguistic practices—we are positioned to start asking modal questions about 'this guy', or whatever our object of reference may be. Without needing to worry about how we have done it, we see ourselves latched onto an object of reference, and ready to ask what its possibilities are. And if the semantic structure of referring expressions is as I shall suggest, we shall find necessary a posteriori truths. If I am right, these truths depend, for their modal status, on our conventions, on conditions we "in default" associate with the expressions. But there is no reason *within* the inquiry for it ever to occur to us that this might be the case. Once we are onto the rigid/nonrigid distinction, our attention is firmly focused upon *objects,* not language.

Thus, even if our referring expressions have the sort of semantic structure I suggest, it is perfectly simple and natural for us to go from the initial presentation of the rigid/nonrigid distinction all the way through to (1)–(5) without ever having to reflect upon our ontology, to address explicitly the question of restrictivism, privileged ontology, and permissivism. But since in our ordinary practice we speak "as if" we accepted a privileged ontology (whether we do or not), it is no big surprise that we can wind up in the position I have urged a defender of (1)–(5) is in—that of *really* being committed to a privileged ontology. The "as if" character of the linguistic practices combines with the absence of metaphysical flags raised during the investigation, so we in effect *assume* a privileged ontology in our interpretation of rigidity, and feel unproblematically determinately focused on "usual" modally extended entities.[20]

Behaving and speaking as if something is true may not amount to assuming or presupposing it. But when this is used—without arguments for its truth—in arguments which explicitly commit us to the view in question, it may more fairly be said that we have presupposed, or accepted a commitment to it *in the course of the argument* —and not just as a *result* of our conclusion. This is what I mean by saying that a privileged ontology is assumed or presupposed in the acceptance of MRD and (1)–(5), and why I am inclined to say that a defense of a privileged ontology is required not only because it is a commitment of MRD and (1)–(5), but because its prior (tacit) acceptance led to these positions.

---

[20] Of course, some may explicitly accept a privileged ontology beforehand. But the ontology still drives the metaphysics, and requires independent defense for the new metaphysics to stand.

The new metaphysics has "seemed" to arise out of mostly linguistic considerations. But it has in fact arisen out of prior acceptance of a privileged ontology disguised in linguistic garb. This metaphysical view needs to be evaluated against its competitors with its own face. Privileged ontology is not a commitment of the new metaphysics—it *is* the new metaphysics. And whatever arguments may be available in their support, the phenomena of rigidity are not among them.

VI SKETCH OF A SEMANTIC ACCOUNT OF RIGIDITY

I conclude with a sketch of an ontologically neutral account of rigidity, which I shall call a *semantic* account, to contrast it with the metaphysical account we have been considering. It is semantic in that counterfactual reference and rigidity (or not) are determined, more or less (see fn. 26), by conditions analytically associated with referring expressions; so if this account is correct, MRD is false, undermining (1)–(5), and the metaphysical view expressed therein. Aside from seeing how these phenomena can occur within nonprivileged ontologies, I hope it will be welcome just to see a semantic treatment presented, as the metaphysical one is usually just assumed.

We begin by considering how there could be determinacy of rigid reference with a permissive ontology, starting again with descriptions. Earlier, we saw that on such an ontology, we would need to associate criteria of identity with 'The President of the U.S.' for it to determinately refer to either Bush or Prez, since its "surface" content applies equally to both. If we also allow such associated criteria to determine the reference of the description in other worlds, then we have both determinate actual reference, and rigidity as well. For the description then applies, in any world, to whatever satisfies the criterion of identity—and, so, to that which is identical to the actual referent—in that world. Determinacy and rigidity are handled in one fell swoop. As we noted, this has the consequence that counterfactual reference is determined by semantic conditions associated with the term, not just by 'identity with the actual referent'—for we do not *have* an actual referent until we have settled on identity conditions. But we are not here concerned to preserve MRD; indeed, our semantic account is supposed to highlight how MRD can fail to be the correct account of rigidity, and how it does fail in nonprivileged ontologies.

One might think that, if we fail to associate the criterion of identity for persons with 'The President of the U.S.', the description is not rigid, but refers nonrigidly to Bush. This, however, ignores the fact that, on our specification, 'being President' is associated as a criterion of *identity* (not just of application), so the referent is *Prez,* even in worlds where Libby Dole is the President (and so we have rigidity). A more serious worry is that it may now seem impossible—as we earlier worried it might be—to have *nonrigid* descriptions. If determinacy of reference requires that we associate a criterion of identity with a description, and this criterion determines its counterfactual application, the referents will always be identical, so all descriptions will be rigid. The only question is *which* entity it rigidly designates.

The supposition that an associated criterion of identity must also serve as the counterfactual criterion of application for a term is mistaken. While this is typical as such matters are usually discussed, there is no reason why we might not associate, with a single expression, one condition as the criterion of identity, and a different one as its criterion of counterfactual application. The two, after all, perform different functions. Criteria of identity tell us under what conditions we have the same object as the actual referent. Criteria of application tell us under what conditions the expression in question applies in another world. I suggest that we need to distinguish the two carefully, and that descriptions (and ultimately, all referring expressions) need to be associated with both criteria of application for other worlds and criteria of identity.

Criteria of identity serve not (of themselves) to determine counterfactual reference, but *actual* reference (by distinguishing among the candidates satisfying the "surface" criteria of application, e.g., 'is the President'); criteria of *application* determine counterfactual reference. When these are the same, as we have seen, the descriptions will be rigid—even when the criterion of identity is "unusual."[21] But the criteria need not be the same; we can associate the identity conditions for people with 'The President'—making Bush the actual referent—while using the familiar descriptive content as our criterion of counterfactual application. This results in—and is, I suggest, how we get—nonrigid descriptions: that to which the expression applies in a world can differ from that which satisfies the criterion of identity.[22]

---

[21] Actually, it is enough if the criteria entail each other. This may be a better picture of the rigidity of complex expressions for numbers.

The associated criteria need not be explicitly formulated. Typically, we recover them from the sorts of counterfactual situations one counts for or against one's statements using these descriptions. As we may postulate an intention to refer rigidly, and so, at least implicitly, through identity, when someone does not count the possibility of a Libby Dole election against their utterance of 'The President could not have been a woman', so we may postulate a more detailed intention if this is needed to get the reference of their expression to be Bush, rather than Prez. Similarly, our judgments of nonrigidity show that we intend the counterfactual application of the term to be given by conditions other than the criterion of identity (though criteria of identity still need to be associated to make the actual referent Bush rather than Prez), and license a similar postulation.

Here is a representation of the semantics of descriptions ('CA' = criterion of application; 'CI' = criterion of identity; '@' stands for the actual world; 'w_x' for other worlds; '(*v*)   ' stands for the value of conditions '   ') [this is a first approximation]:

(1) Rigid descriptions (normal) (e.g., 'The President' rigidly designating Bush):[23]
    CA@ = Being the President (one might spell this out further)
    CAw_x = (say) Having the same origin (in w_x) as (*v*)CA@
    CI = Having the same origin as (*v*)CA@

(2) Rigid descriptions (unusual) (e.g., 'The President' rigidly designating Prez)
    CA@ = Being the President
    CAw_x = Being the President
    CI = Being the President[24]

(3) Nonrigid descriptions ('The President' refers nonrigidly to Bush)
    CA@ = Being the President
    CAw_x = Being the President
    CI = (say) Having the same origin as (*v*)CA@

---

[22] The question of why we might ever give these criteria different conditions is an interesting one, and there are interesting things to be said about why in practice philosophers tend to run them together, but space limitations preclude a fuller discussion here.

[23] If (as I think) Keith Donnellan's distinction is genuinely semantic, the actual criteria for "referential" descriptions would be given not by the "straight" content of the description, but by causal conditions (see the account of names below), or some other "intended" content.

[24] I realize this does not really have the normal form of a criterion of identity—'*x* bears *R* to *y*'. The sort of criteria of identity entities like Prez have allow them to be specified nonrelationally.

The "surface" sense of a description gives the parameters of the actual conditions of application, the criteria of application for other worlds say what conditions something must meet in another world to be the referent of the expression in that world, and the criteria of identity specify the conditions something in another world must meet to be identical to the actual referent—and thus also determine the actual referent.[25] Note that on the usual treatment, we would fill in 'CI', for rigid descriptions, with "Being identical to the actual value of the description." But in a permissive ontology, this will not suffice—for it is not determined whether it is identity with Bush or Prez that is indicated. The identity conditions must at least be specified according to the ***kind of thing*** in question, if not more finely.[26]

In normal rigid descriptions, we let the criteria of identity serve to determine the counterfactual criteria of application. This is the kernel of truth, in a permissive ontology, in the idea that we can refer through identity: we can refer through *criteria of identity*, so it is "built in" semantically that the expression has the same reference in all worlds. To this extent, descriptions can be rigid *de jure* (see fn. 13).[27] But since descriptive conditions, rather than the metaphysics of identity determine the reference, MRD fails here.

This, in brief sketch, gives our semantic account of rigidity. The rules of reference, for both rigid and nonrigid descriptions, are given semantically, not metaphysically. Because of this, the account can also work within a restrictive ontology. In this case, one might say that it is not the metaphysics of identity which determine the counterfactual referents of rigid terms, but rather, the semantics of rigidity which determine what is to count as identical across worlds. And while in a *privileged* ontology, rigidity can be determined metaphysically, there is nothing to rule out the semantic story being just as I have outlined it. Thus, the semantic account of rigid designation is neutral with respect to our three "worldviews," while the "metaphysical" account can work only within a privileged ontology. This, I believe, provides some ground to prefer my treatment whatever one's metaphysics—for while the metaphysical questions are matters for dispute, it hardly seems a real matter of dispute whether we can and do use expressions rigidly.

---

[25] Criteria of counterfactual application mentioning (*v*)CA@ do not quite just plug in the actual value of the descriptions (for the CA@ do not of themselves determine an actual value), but are more like functions from the "actual" content *and* identity conditions to objects in other worlds—and the criteria for the actual world may be seen either as providing ambiguous actual reference, or better (at least usually), as functions from "surface" content *and* criteria of identity to a referent. This is why the above sketch is a "first approximation." Criteria of identity could thus be considered as part of the criteria of actual application, but it is useful to distinguish their contributions. Note that criteria of identity play the same role in getting determinate reference in other worlds. In theory, there is room for distinct criteria of identity for the actual and counterfactual referents (when $\mathrm{CA}w_n$ is not given by CI). But I am not aware of any expressions that actually function this way

[26] I mean here to reemphasize that the criteria need not fully specify the features counterfactual referents of rigid terms must have (see fn. 7)—(1) above leaves open what the origin is, and so what it is in other worlds (though not that it is the same in all worlds)—thus allowing for necessary a posteriori truths ('Bush's origin is   '). But it would be misleading to speak of this as the empirical discovery of essence, for what is empirical is just the value of parameters, the essentiality of which is set by the language, and a priori.

[27] Within this framework, then, the *de jure/de facto* distinction becomes blurred, or imperspicuous. If we identify rigid *de jure* expressions as those which are descriptionless, as in MRD, then there are no such expressions. If we require only that rigidity be semantically built in, there is really no rigidity *de facto*. Perhaps the intuitive distinction divides terms where the criteria are the same from those which imply each other.

Finally, back to names. For a name to be rigid, it must be associated with a criterion of identity across worlds, and this then serves as the criterion of application for the name, viz:

(4) Names ('Bush'):
CA@ = Description/ostension/causal relation
CAw_n = Whatever satisfies CI in w_n
CI = (say) Having the same origin as (*v*)CA@

Names, then, like rigid descriptions, are rigid not by lacking descriptive content, but because their criteria of application are the same as (given by) their associated criteria of identity. This allows us, in theory anyway, to identify a possible nonrigid use of names in which these two are separated ('Superman' or 'Mr. President' are possible examples). We may think of such names as disguised descriptions; the point is just that nothing prevents "ordinary names"—linguistic items with no surface syntactic or semantic structure—being nonrigid. We simply do not ordinarily so use them.

To restate my main contention: once we distinguish the phenomenon of rigidity from its metaphysical interpretation, and see that the latter hinges on just the metaphysical view the phenomena are thought to support, we can refocus the debate at the metaphysical level, where it belongs. The "new essentialism" is nothing more nor less than advocacy of a privileged ontology, and needs to be defended as such—as a metaphysical view that neither flows from nor is supported by the fact that some, but not all expressions are rigid. By now, it is probably clear that I am rather skeptical about the possibility of a successful defense of this view. But arguments await another occasion.

ALAN SIDELLE

University of Wisconsin/Madison

A POPULAR PRESUMPTION REFUTED

Compare the single-event interpretations of

(1) Sly Pete will win.
(2) Sly Pete wins.
(3) Sly Pete won.

Call them m_1, m_2, and m_3, respectively. It is popularly presumed that these three differ only in the times they assign to Sly Pete's win, and more particularly that m_1 formulates about the future exactly what m_2 and m_3 formulate about the present and past—whatever that might be. I shall demonstrate that this presumption cannot be right.

My argument is premised on a theory and a fact. The theory is about how reported speech works. The fact is about what English speakers say to report an affirmation of m_1—or rather what they do not say. To keep the two premises separate, I shall use different examples when doing the theory, beginning with the single-state interpretation m_4 of 'I am awake'.

If a speaker *X* affirms m_4 on Monday, then as a matter of observation a different English speaker *Y* will report this on Tuesday by saying '*X* said she was awake'. The single-state interpretation m_5 of the embedded sentence 'she was awake' is of course a different message from m_4, the latter attributing present wakefulness to whoever the speaker is, the former past wakefulness to a third person. But each provides its speaker with means of referring to the same matter of historical fact. 'I' is *X*'s way, and 'she' *Y*'s, of identifying *X* as the person whose wakefulness at *X*'s moment of utterance is located as present by m_4 and past by m_5. That then is how reported speech is meant to work: the aim is to preserve objective substance in the face of changed circumstances of transmission.[1]

Taken together with the presumption that m_1 merely formulates about the future what m_3 formulates about the past, this theory entails that, if *X* affirms m_1 regarding the coming evening, *Y* ought to report this next day by saying '*X* said *Sly Pete won.*' For with the relevant evening future with respect to *X*'s point of speech and past with respect to *Y*'s, m_3 achieves exactly the objective substance for *Y* that m_1 achieved ex hypothesi for *X*. But as a matter of observation this is not so. Therefore the presumption is false; q.e.d.

---

[1] In particular, location in the *A*-series is adjusted to preserve location in the *B*-series. See J.M.E. McTaggart, "The Unreality of Time," *Mind*, XVII (1908): 457–84, esp. 458. This is all that matters for my demonstration

0022-362X/92/8908/431–2

If this argument works at all it works for any natural language. For instance, we are standardly taught that 'vincera' and 'vinse' are, respectively, a "future-tense" and a past-tense form of the Italian verb to win. But if the guiding principle of Italian reported speech is to preserve objective substance—an issue concerning which the evidence is there to be culled—then we must resist the natural implication that the only difference in meaning between the two forms concerns the time of the win, because as a matter of observation 'vincera' is reported after the event by something quite other than 'vinse', namely, 'avrebbe vinto'.

The argument applies in narrower cases as well. Consider the single-events interpretations m_6 and m_7 of

(6) If Sly Pete calls he will win.
(7) If Sly Pete called he won.

Observation of how English speakers report prior affirmations of m_6 reveals the impossibility that m_6 might merely formulate about the future what m_7 formulates about the past. Yet every theory of "conditionals" I know seems to presume exactly that, recommending the very same semantic treatment of each.

A swifter refutation. What *Y* in fact says is '*X* said Sly Pete *would* win', or '*X* said that if Sly Pete called he *would* win'.[2] But according to our popular presumption, the function of 'will' in (1) is to locate Sly Pete's win directly in the future, just as 'won' in (3) locates it directly in the past. And nothing 'would' could possibly do could match the sheer simplicity of this arrangement. The popular presumption leaves it inexplicable why *Y* should resort to 'would' to report *X*'s 'will'.

V. H. DUDMAN

Macquarie University

---

[2] Elsewhere I have advanced a theory that can explain why this should be so—most recently in parts 4 and 5 of "Interpretations of 'If'-sentences," in *Conditionals*, Frank Jackson, ed. (New York: Oxford, 1991), pp. 202–32. But here my purpose is wholly negative.

BOOK REVIEWS

Wittgenstein's Philosophy of Psychology. MALCOLM BUDD. New York: Routledge, 1989. xiv + 186 p. Cloth $39.95, paper $14.95.

Sapiently gauging his distance, Malcolm Budd offers us a synoptic view of Wittgenstein's later thoughts on the mind. His treatment is selective yet representative. Seven dense chapters survey Wittgenstein's various surveys of meaning and understanding, sensations and sense impressions, the seeing of aspects, images, internal speech and calculation, thought and intention, and feelings, emotions, and bodily awareness. Budd's expositions are in every way exemplary. He is deeply sympathetic to his subject but by no means uncritical. I know of no other work on Wittgenstein that can match his for clarity, accuracy, concision, and penetration (and there are some fine books out there). He knows *exactly* what he is talking about and is able to convey this to the reader with unostentatious efficiency. His book would be perfect for students still trying to find their feet with Wittgenstein, but it can be heartily recommended to scholars who think they know their way around the Austrian's later thought. Here is a book on Wittgenstein that you can *trust.*

Beginning with a chapter on method, Budd emphasizes the prophylactic character of Wittgenstein's intentions: the notions of grammar, language game, and perspicuous representation are to be seen in a curative light. He brings out the consequent radical disjuncture Wittgenstein perceived between our ordinary mastery of psychological concepts and our bewilderment once we reflect upon this mastery: we go from genius to dolt simply by taking a downward glance at our concepts; we simultaneously grasp and do not grasp what we mean. Budd is sensitive to the apparent paradox this presents and does his best to explain how Wittgenstein lived with it. I feel that this issue could stand further work: Did Wittgenstein really have an adequate account of the depth of error he attributed to our reflective understanding, given that he took our ground-floor conceptual grasp to be essentially flawless?

Chapter II sets out Wittgenstein's position on meaning, rules, conscious processes, mental mechanisms, capacities, reasons, and communication: These topics have been hotly contested in recent philosophy, stimulated largely by Saul Kripke's "community interpretation" of Wittgenstein's position. Budd shows definitively, I think, that this misrepresents the essence of Wittgenstein's view; he stresses instead the contrast between inner and outer. The main

0022-362X/92/8908/433–6

target of Wittgenstein's critique is the idea that meaning is a quality of consciousness—something that happens *in* you.

Wittgenstein is surely right about this, but it does not follow that the concepts of meaning and consciousness are unconnected, nor that the latter is not a necessary condition of the former. Meaning can be (and be essentially) something we do consciously without thereby consisting in an introspectible quality of consciousness. Wittgenstein seems not to have explored other less direct ways in which meaning and consciousness might intersect.

Budd's presentation of the private-language argument is impressively clear and thorough. Since first-person ascriptions are criterionless, private sensations could not afford the constraint on correct use which governs our actual practice of ascribing behaviorally expressed sensations. The would-be private linguist is thus bereft of any means of establishing a reference relation. Budd appears to accept this argument, while plausibly criticizing Wittgenstein for extruding causality from his account of sensation concepts. I would say that the introduction of a causal relation between sensations and self-ascriptions provides one way of resisting the argument. We should, first, remind ourselves of how surprising the conclusion of Wittgenstein's argument ought to seem. He invites us to believe that someone whose sensations happen to have no behavioral expression is semantically impotent in respect of those sensations. The subject has the sensations, she can speak a language for public objects, but somehow she just cannot get her words to refer to her sensations. Only if her body offers up potential criteria for others to detect her sensations can she succeed in homing in semantically on those sensations. Suppose the sensations are publicly undetectable up to time *t* and then acquire behavioral manifestations at *t*, only to revert to privacy ten minutes later. According to Wittgenstein, the subject could not refer to her sensations before *t*, though she can once her body starts to show their presence to others; and when the behavioral manifestations go her sensation words lapse back into mere empty sounds. Surely this is a remarkable result: it ties my semantic abilities with respect to my own mental life to the epistemic powers of others in detecting what it is I am experiencing.

The suspicion that some form of verificationism underlies Wittgenstein's argument is reinforced if we consider what a causal theory of reference might have to say about the possibility of a private lexicon, since such theories do not locate reference in the speaker's ability to *check* that he is referring correctly—only in the objective obtaining of causal relations between objects and uses (think, say, of Jerry Fodor's asymmetric dependence theory). Since Budd accepts that sensations cause self-ascriptions, he is not precluded from considering a theory of this kind. And if a causal theory can be applied to sensation words, we have this result: so long as the private objects bear the appropriate reference-conferring causal relations to first-person uses, that will be sufficient to underpin a semantic relation—no matter *how* private the sensations are. In other words, causal theories impose reference conditions that are inclusive enough to permit a private language (the same could be said of teleological theories). For such theories do not build in any necessity for anyone to be able to *verify* that words are being used as their semantics demands; it simply has to *be* the case that tokens of 'pain' (say) appropriately covary with the occurrence of pains in the speaker. These theories may or may not be adequate, but we need to consider them if we are interested in whether the privacy of a sensation thwarts reference to that sensation.

Nor should we let the general mysteriousness of the reference relation induce us to suppose that reference is uniquely impossible where private objects are concerned, since that would be to pin the blame on the wrong thing. Maybe we cannot really explain reference in *any* context; it is just that this becomes more obvious when we strip reference down to its purest form. In any event, the private-language argument seems to me a good deal more questionable than it seems to Budd.

The chapter on aspect perception is a subtle and lucid discussion of the way in which Wittgenstein demonstrates the "polymorphous character" of the concept of seeing. Seeing an aspect is neither quite like seeing color and shape, but nor is it a case of interpreting what one strictly sees. The fact is that it lies between sensation and thought and cannot be assimilated to either. The phenomenon thus serves to expose the simplifying errors inherent in both empiricist and rationalist views of mind. Our mental concepts are much less monolithic than philosophers have supposed.

On the distinction between visualizing and seeing, Budd ably expounds Wittgenstein's idea that the difference consists essentially in subjection to the will. One's attitude in visualizing is active, while in seeing one is a passive observer. The difference is not a matter of experiential content or quality. This is held, plausibly, to explain what may appear anomalous: that one cannot see and visualize the same thing at the same time, though one can do this for different things—since one cannot be active and passive toward the same experiential content simultaneously.

Wittgenstein's opposition to the idea of an inner vehicle of thought is shown to be an instance of his wider point that nothing of an intrinsically nonrepresentational nature could ever be the basis of one's knowledge of the content of one's thought, so that the putative inner vehicle could play no epistemic role in an account of self-knowledge of intentional states. I would agree with this point but note that it does not follow that the inner vehicle might not be needed for other purposes. Indeed, given Budd's earlier insistence on the causal properties of sensations, he might regard with sympathy contemporary thinkers (like Fodor) who maintain that no account of intentional causation is possible without introducing an inner vehicle whose structure encodes the causal powers of thought.

The book concludes with a fresh discussion of the relation between emotion and proprioception, on the one hand, and somatic sensation, on the other. As we might expect by now, Wittgenstein's main target is the idea that bodily feelings are constitutive of emotion or proprioception. Such feelings *accompany* one's fear or one's awareness of the position of one's arm, but the concepts are not concepts *of* such feelings—rather as certain sensations accompany linguistic understanding without constituting it. Nor do we know our emotional state or bodily position by means of our awareness of the accompanying sensations. This explains something that might otherwise seem puzzling: how a horrible or agreeable emotion can be so much more powerful than the somatic perturbations that go with it. Affective tone is more a matter of the intentionality of the emotion than its bodily feel.

In the preface to his book, Budd makes a nice acknowledgement to me, adding that "he has not seen what I have written and is unlikely to agree with all of it." Well, now I have seen it and agree with far more of it than might be thought proper in a reviewer (and not merely because we are friends). The book is a model of Wittgenstein exegesis. Only on the soundness of the private-language argument do we seriously disagree: he thinks it works and I do not.

COLIN McGINN

Rutgers University

Locke: Volume I, Epistemology; Volume II, Ontology. MICHAEL AYERS. New York: Routledge, 1991. x + 341 p. and vi + 341 p. Cloth $127.50 (set).

This book is something of a *tour de force.* If, as I understand, it has been a long time in the writing, the time was well worth taking. Over

0022-362X/92/8908/436–40

its very considerable length it sustains (and deserves in response) an unflaggingly high level of serious thought. The bare simplicity of its title belies its complexity and richness. In an introduction that discusses the relation of philosophy to its history, Ayers says that, being intended primarily for philosophers, the book "is probably not historical enough" in its *purposes* to suit intellectual historians, but that the *result* "is probably too historical to please many philosophers" (I, 3–4). This need not mean that neither historians nor philosophers could be satisfied by it, for besides the chapters on Locke's philosophical antecedents and surroundings (very useful to the student of *any* seventeenth-century philosopher), and those on details of Locke's own arguments, there are long stretches of pure and unadulterated philosophy. What it unfortunately may mean is that the material of this latter kind may not only be passed over by intellectual historians, but also missed by philosophers who might never dream of picking up a book called "Locke." Although it manages nicely to combine an analytic approach to "the arguments of the philosophers" (the title of the series for which it was written) with some very illuminating history, much of the philosophy in it is stimulated by, rather than done along with Locke.

Ayers's aim is to "recover the meaning" (II, 1) of Locke's *Essay*, and to assess what emerges by comparison with relevant historical rivals and also "current" theories. It is wry understatement to say that this comparison is "not always to the advantage of the present" (II, 1). A very valuable, and certainly one of the most characteristic and striking things about the book, is the extent to and manner in which Ayers comes to the defense of Locke. This is not a simple matter of straightforward sympathy as might be expressed in an attempt to explain, from an historically enlightened perspective, how Locke could have come to say things that might now seem patently unacceptable. It is, rather, a matter of Ayers's believing (and this is part of his defense of doing history of philosophy at all) that we can learn from the past—not, let it be stressed, by way of identifying mistakes to avoid repeating, but by way of recovering lost truths. "The past is an invaluable repository of unfashionable truth for those with the sympathy, judgement and patience to discern and make use of it" (I, 9).

In effect, each of the two volumes has as its theme a truth of which recent philosophy has lost sight, and which the study of Locke helps us to regain. A continuing theme of the first volume is the rightness of Locke's theory of knowledge to ascribe a central role to subjectivity or consciousness. Along with what he finds in Locke, Ayers argues that perceptual beliefs have to be understood as being such that their source and objects are present to consciousness in a direct and perspicuous manner. This is not to say that they are infallible, and any suggestion of Locke's that they are "evident" needs carefully to be detached from this. Yet it is, says Ayers, a perfectly reasonable rejection of the idea of guaranteed knowledge deliverance which is partly responsible for subjectivity's being left out of modern accounts of knowledge. "The chief, and for some the only, argument for accepting the phenomenon of consciousness has been its obvious existence, rather than its indespensability for explaining the possibility of knowledge" (I, 193). The idea of primary and direct knowledge that has an "evidence" or "perspicuousness" leads to a defense of "a qualified form" (I, 268) of Locke's intuitionism about a priori knowledge, with its talk of "seeing" necessary connections between ideas. It leads also to a criticism of twentieth-century "linguistic" accounts of such knowledge.

The continuing theme of the second volume (which has already made itself felt in the background in the first) is a wholehearted and energetic rejection of "linguistic idealism" or "conceptual relativism." Ayers argues that language- and concept-independent reality itself "gives" us distinct objects. He argues for a strong and objective contrast, of a sort which, he says, Locke "accepted, stated, and defended" between "substances, which are the real and natural unities, and non-substances which are in some sense abstractions or creatures of the mind" (II, 10). One the treasures of this book is Ayers's long and detailed, intellectually very rich and sensitive, exploration and development of this distinction between "substances" and "modes."

None of this is to say that Locke got it all right. What Ayers develops and defends are theses "more or less analogous" (II, 2) to Locke's. Locke wrongly assumed that the perspicuity of primary knowledge provides a criterion of truth, and the twentieth century is right to reject that assumption. What is needed is "authoritative perpiciousness without infallibility" (II, 3). Moreover, Locke's causal theory of representation robs experience of any intentionality. Modern thought is wrong to approve this in Locke, wrong not to see that it is conscious thought, and not language, or behavior, or functional states of the brain, which is the bearer of intentionality. For Locke, sensations are "blank effects" that represent things only in virtue of being caused by them; whereas, argues Ayers, they are conscious states which have an intentional content. Only the "*intrinsic* intentionality of sensation, can explain the essential epistemological role of sense-experience or make it seem anything but an accident that sensations occur in the process which gives rise to perceptual knowledge" (I, 172). Furthermore, Locke's quasi-perceptual intuitionism about necessary truth fails to explain our apprehension of modality.

As for the objective distinction between the categories of substance and mode, Locke does not always deal with it properly. So, far from anticipating Saul Kripke, his account of their names wrongly supposes that the belief that substances have real essences is "semantically inert" (II, 78). He failed also in not giving due theoretical weight to the category of "natural modes"—a category that is produced by a natural/nonnatural cut across the substance/mode distinction. (This category includes natural but nonsubstantial things, such as measles or photosynthesis, as opposed to Locke's nonnatural modes, such as triangle or property; and it is one about which Ayers is most interesting.) This latter failure of Locke's stems from his trying to account for the "natural unity" of substances as opposed to the mind-given unity of modes, in terms of a principle of causality—a unifying causality which, evidently, is there with natural modes, too. Though "not entirely mistaken" (II, 128), this causal unity is more superficial than what is required, namely, a conception of *material* unity. We need, Ayers argues, a serious idea of the material coherence of a lump of stuff—an idea that Locke conflated with the different idea of quantity of stuff with the result that, along with more recent philosophy, he wrongly supposed that entities could be constructed by fiat by the arbitrary addition of principles of identity. Against their downgrading and underestimation by philosophers of all ages, Ayers defends materially unified chunks, such as rocks, as crucially important ontological items.

One feature of the way in which, having recovered his meaning, Ayers wishes to assess Locke, is to take his arguments "not just one by one, but as a whole," and to make a response which is as "systematic as the philosophy under consideration" (I, 5–6). He does not, then, attempt to repeat or explain, either in detail or in general, the surface structure of the *Essay* as it presents itself to someone less knowledgeable than Ayers. From the very outset, he clearly assumes his readers will already be rather familiar with the text. The discussion is usually so far below its surface as to be rather unsystematic in terms of the *Essay*'s books, chapters, and sections. A somewhat startling symptom of this is that the last ten or twelve percent of the whole is devoted to Locke's chapter on identity. This is a chapter that was unknown in the *Essay* as first published, and it is, Ayers has to concede, in some respects "an awkward afterthought." But in other respects, he says, it is "a natural culmination of Locke's system" (II, 205); and his discussion of it does indeed come, not merely at the end of, but as a climax to, the whole work. It has to be said that in the light of Ayers's endeavors to read Locke "systematically" one can see how it would be possible to be a "Lockean" in the way it always has been to be a "Kantian."

Throughout the book there is little dialogue with recent writers on Locke. On more than one occasion, "some commentators" are mentioned without being identified or located. No doubt this has something to do with the fact that Ayers is not taking Locke's arguments "one by one." But it unfortunately harmonizes with the sense of a mannered disdain which is produced by the provocative loftiness of the all too frequent dismissive references to the "fashionable," "the currently popular," "the latest thing." These references go from being annoying to being ludicrously parochial when, having found someone whose thought has recently converged with his, Ayers comments, on the very last page: "it must be the Oxford air, although not every argument out of Oxford is as cogent as this one seems to be" (II, 238).

R. S. WOOLHOUSE
University of York

THE JOURNAL OF PHILOSOPHY
is pleased to announce the publication of its
TWENTY-FIVE YEAR INDEX, 1964–1988
VOLUME LXI TO VOLUME LXXXV
AUTHORS, SUBJECTS, & REVIEWS
Supplement to its
FIFTY-YEAR INDEX, 1904–1953
TEN-YEAR SUPPLEMENT, 1954–1963

It is now available at special prices
25-year (1964–1988) $25.00 cloth/$10.00 paper
also available
50-year (1904–1953) $12.00 cloth 10-year (1954–1963) $3.50 paper
Shipping $5.00 (first)/$2.50 (each addl.)

Please send a check or money order to

Journal of Philosophy
704-I Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through May 31, 1993. Orders and payment must be received by then. Paper copies will be mailed immediately, but cloth copies will take 6–8 weeks for delivery.

NEW BOOKS: PAPERBACKS

ARISTOTLE: *On Rhetoric.* GEORGE A. KENNEDY, trans. and ed. New York: Oxford, 1992. xvi, 335 p. $9.95.

BEHLER, ERNST: *Confrontations: Derrida/Heidegger/Nietzsche.* STEVEN TANBENECK, trans. Stanford: University Press, 1992. xii, 184 p. $11.95.

BERNAUER, JAMES W.: *Michel Foucault's Force of Flight: Toward an Ethics for Thought.* Atlantic Highlands, NJ: Humanities, 1992. xii, 261 p. $15.00.

CHURCHLAND, PAUL M.: *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge: MIT, 1992. xvi, 321 p. $15.95.

FALCK, COLIN: *Myth, Truth, and Literature: Towards a True Post-Modernism.* New York: Cambridge, 1992. xv, 173 p. $15.95.

GRIFFITHS, A. PHILLIPS, ed.: *A. J. Ayer: Memorial Essays.* New York: Cambridge, 1992. v, 239 p. $18.95.

KIERKEGAARD, SØREN: *The Concept of Irony.* HOWARD V. HONG and EDNA H. HONG, eds. and trans. Princeton: University Press, 1992. xxv, 634 p. $19.95.

MADELL, GEOFFREY: *Mind and Materialism.* Edinburgh: University Press, 1991. vii, 151 p. $18.00.

MARGOLIS, JOSEPH: *The Truth about Relativism.* Cambridge: Blackwell, 1992. xvi, 224 p. $19.95.

MARTIN, MICHAEL: *Atheism: A Philosophical Justification.* Philadelphia: Temple, 1992. xiii, 539 p. $24.95

NANCY, JEAN-LUC: *The Inoperative Community.* PETER CONNOR, ed. and trans. Minneapolis: Minnesota UP, 1991. xli, 176 p. $14.95.

NIETZSCHE, FRIEDRICH: *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s.* DANIEL BREAZEALE, ed. and trans. Atlantic Highlands, NJ: Humanities, 1992. lxv, 165 p. $15.00.

PARMENIDES OF ELEA: *Fragments: A Text and Translation.* DAVID GALLOP, trans. Toronto: University Press, 1991. x, 144 p. $17.95.

PUTNAM, HILARY: *Representation and Reality.* Cambridge, MIT, 1991. xv, 136 p. $9.95.

WURZER, WILHELM S.: *Filming and Judgement: Between Heidegger and Adorno.* Atlantic Highlands, NJ: Humanities, 1992. xviii, 149 p. $12.50.

YOVEL, YIRMIYAHU: *Spinoza and Other Heretics: The Adventures of Immanence.* Princeton: University Press, 1992. xv, 231 p. $12.95. And *Spinoza and Other Heretics: The Marrano of Reason.* Princeton: University Press, 1992. xiii, 248 p. $12.95.

NEW BOOKS: TRANSLATIONS

ADORNO, THEODOR W.: *Notes to Literature, Vol. II.* SHERRY WEBER NICHOLSEN, trans New York: Columbia, 1992. xx, 350 p. Cloth $35.00.

AGAMBEN, GIORGIO: *Language and Death: The Place of Negativity.* KAREN E. PINKUS, trans. Minneapolis: Minnesota UP, 1991. xiii, 112 p. Cloth $34.95, paper $14.95.

APHRODISIAS, ALEXANDER OF: *Quaestiones 1.1-2.15.* R. W. SHARPLES, trans. Ithaca: Cornell, 1992. 183 p. Cloth $47.95. And *On Aristotle's* Metaphysics *2 and 3.* WILLIAM E. DOOLEY and ARTHUR MADIGAN, trans. Ithaca: Cornell, 1992. 242 p. Cloth $47.95.

BENITEZ-ROJO, ANTONIO: *The Repeating Island: The Caribbean and the Postmodern Perspective.* JAMES E. MARANISS, trans. Durham: Duke, 1992. xi, 302 p. Cloth $49.95, paper $15.95.

CROCE, BENEDETTO: *The Aesthetic as the Science of Expression and of the Linguistic in General.* COLIN LYAS, trans. New York: Cambridge, 1992 xxxiv, 172 p. Cloth $49.95, paper $16.95.

DELEUZE, GILLES: *Expressionism in Philosophy: Spinoza.* MARTIN JOUGHIN, trans. Cambridge: MIT, 1992. 445 p. Cloth $16.95.

DERRIDA, JACQUES: *Cinders.* NED LUKACHER, trans. Lincoln: Nebraska UP, 1992. 79 p. Cloth $25.00. And *The Other Heading: Reflections on Today's Europe.* PASCALE-ANNE BRAULT and MICHAEL B. NAAS, trans. Bloomington: Indiana UP, 1992. lix, 129 p. Cloth $19.95.

DUHEM, PIERRE. *German Science.* JOHN LYON, trans. LaSalle, IL: Open Court, 1991. xxv, 136 p. Cloth $38.95, paper $17.95.

FERRY, LUC: *Political Philosophy, Vol. II: The System of Philosophies of History.* FRANKLIN PHILIP, trans. Chicago: University Press, 1992. x, 200 p. Cloth $26.95.

HEIDEGGER, MARTIN: *The Concept of Time: English-German Edition.* WILLIAM MCNEILL, trans. Cambridge: Blackwell, 1992. viii, 40 p. Cloth $34.95, paper $14.95. And *Parmenides.* ANDRE SCHUWER and RICHARD ROJCEWICZ, trans. Bloomington: Indiana UP, 1992. xv, 170 p. Cloth $29.95. And *The Principle of Reason.* REGINALD LILLY, trans. Bloomington: Indiana UP, 1991. xxi, 148 p. Cloth $27.50.

HORSTER, DETLEF: *Habermas: An Introduction.* HEIDI THOMPSON, trans. Philadelphia: Pennbridge, 1992. viii, 110 p Paper $9.95.

LOONE, EERO: *Soviet Marxism and Analytical Philosophies of History.* BRIAN PEARCE, trans. New York: Verso, 1992. xvii, 280 p. Cloth $69.95.

LYOTARD, JEAN-FRANCOIS: *The Inhuman.* GEOFFREY BENNINGTON and RACHEL BOWLBY, trans. Stanford: University Press, 1992. viii, 216 p. Cloth $37.50.

NISHIDA, KITARO: *An Inquiry into the Good.* MASAO ABE and CHRISTOPHER IVES, trans. New Haven: Yale, 1992. xxxiv, 184 p. Cloth $25.00, paper $12.00.

SIMPLICIUS: *Corollaries on Place and Time.* J. O. URMSON, trans. Ithaca: Cornell, 1992. 157 p. Cloth $47.95.

VAN REIJEN, WILLEM: *Adorno: An Introduction.* DIETER ENGELBRECHT, trans. Philadelphia Pennbridge, 1992. viii, 92 p. Paper $9.95

NEW BOOKS: ANTHOLOGIES

ALPERSON, PHILIP, ed.: *The Philosophy of the Visual Arts.* New York: Oxford, 1992. x, 630 p. Paper $29.95.

ANNAS, JULIA, ed.: *Oxford Studies in Ancient Philosophy, Supplementary Vol. 1992: Methods of Interpreting Plato and His Dialogues.* New York. Oxford, 1992. 280 p Cloth $65.00. And *Oxford Studies in Ancient Philosophy, Vol. IX.* New York: Oxford, 1992. 221 p. Cloth $62.00.

AVINERI, SHLOMO, and AVNER DE-SHALIT, eds.: *Communitarianism and Individualism.* New York: Oxford, 1992. vi, 237 p. Cloth $56 00, paper $19.95.

AYER, A. J : *The Philosophy of A. J. Ayer.* LEWIS EDWIN HAHN, ed. La Salle, IL: Open Court, 1992. xix, 696 p. Paper $24.95.

BEAKLEY, BRIAN, and PETER LUDLOW, eds.: *The Philosophy of Mind: Classical Problems, Contemporary Issues.* Cambridge: MIT, 1992. ix, 433 p Cloth $39.95, paper $19.95.

BOCKOVER, MARY I., ed.: *Rules, Rituals, and Responsibility: Essays Dedicated to Herbert Fingarette.* La Salle, IL: Open Court, 1992. xxvii, 228 p. Cloth $48.95, paper $21 95.

CAMPBELL, JAMES, ed.: *Selected Writings of James Hayden Tufts.* Carbondale: Southern Illinois UP, 1992. lvi, 433 p. Cloth $45.00.

CAPLAN, ARTHUR L., ROBERT H. BLANK, and JANNA C. MERRICK, eds.: *Compelled Compassion: Government Intervention in the Treatment of Critically Ill Newborns.* Totowa, NJ: Humana, 1992. xi, 336 p. Cloth $39.50.

COLE, EVE BROWNING, and SUSAN COULTRAP-MCQUIN, eds.: *Explorations in Feminist Ethics: Theory and Practice.* Bloomington: Indiana UP, 1992. xi, 203 p. Cloth $35.00, paper $12.95.

CRANE, TIM, ed.: *The Contents of Experience: Essays on Perception.* New York: Cambridge, 1992 ix, 275 p. Cloth $49.95.

DETLEFSEN, MICHAEL, ed.. *Proof and Knowledge in Mathematics.* New York: Routledge, 1992. x, 256 p. Cloth $52.50.

DEUTSCH, ELIOT, ed.: *Culture and Modernity: East-West Philosophic Perspectives.* Honolulu: Hawaii UP, 1992. xvii, 641 p. Cloth $48.00.

FINE, A., M. FORBES, and L. WESSELS, eds.: *PSA 1990: Proceedings of the Philosophy of Science Association Biennial Meeting, Vol. II.* East Lansing, MI: Philosophy of Science Association, 1991. xix, 575 p. Cloth $25.00.

FRAZER, ELIZABETH, JENNIFER HORNSBY, and SABRINA LOVIBOND, eds.· *Ethics: A Feminist Reader.* Cambridge: Blackwell, 1992. xi, 500 p Cloth $49.95, paper $19.95.

FRENCH, PETER A., THEODORE E. UEHLING, JR., and HOWARD K. WETTSTEIN, eds.: *Midwest Studies in Philosophy, Vol. XVI: Philosophy and the Arts.* Notre Dame: University Press, 1992. 496 p. Cloth $49.95, paper $21.95.

FU, CHARLES WEI-HSUN, and SANDRA A. WAWRYTHKO, eds.: *Buddhist Ethics and Modern Society.* Westport, CT: Greenwood, 1991. xxvii, 441 p. Cloth $55.00.

GEORGE, ROBERT P., ed.: *Natural Law Theory: Contemporary Essays.* New York: Oxford, 1992. x, 371 p. Cloth $39.95.

HANFLING, OSWALD, ed.: *Philosophical Aesthetics: An Introduction.* Cambridge: Blackwell, 1992. xxvi, 483 p. Cloth $54.95.

HESTER, MARCUS, ed.: *Faith, Reason, and Skepticism: Essays.* Philadelphia: Temple, 1992. 180 p. Cloth $39.95.

JUDSON, LINDSAY, ed.: *Aristotle's* Physics: *A Collection of Essays.* New York: Oxford, 1992. 286 p. Cloth $65.00.

KROLL, RICHARD, et al., eds.: *Philosophy, Science, and Religion in England 1640–1700.* New York: Cambridge, 1992. xv, 287 p. Cloth $59.95.

LEYH, GREGORY, ed.: *Legal Hermeneutics: History, Theory, and Practice.* Berkeley: California UP, 1992 xix, 325 p. Cloth $45.00, paper $15.95.

LOEWER, BARRY, and GEORGES REY, eds.: *Meaning in Mind: Fodor and His Critics.* Cambridge: Blackwell, 1991. xxxvii, 343 p. Cloth $44.95.

MARANHAO, TULLIO, ed.: *The Interpretation of Dialogue.* Chicago: University Press, 1990. vii, 376 p. Cloth $47.00, paper $17.95.

MEYER, MICHAEL J., and W. A. PARENT, eds.. *The Constitution of Rights: Human Dignity and American Values.* Ithaca: Cornell, 1992. viii, 248 p. Cloth $32.50, paper $12.95.

MILL, JAMES: *Political Writings.* TERENCE BALL, ed. New York: Cambridge, 1992. xxxvii, 317 p. Cloth $49.95, paper $15.95.

MITIAS, MICHAEL H., ed.: *Moral Education and the Liberal Arts.* Westport, CT: Greenwood, 1992. viii, 177 p. Cloth $45.00.

MORGAN, MICHAEL L., ed.: *Classics of Moral and Political Theory.* Indianapolis: Hackett, 1992. viii, 1288 p. Cloth $48.50, paper $27.50.

MOUFFE, CHANTAL, ed.: *Dimensions of Radical Democracy: Pluralism, Citizenship, Community.* New York: Verso, 1992. vii, 254 p. Cloth $64.95, paper $18.95.

NARDIN, TERRY, and DAVID R. MAPEL, eds.: *Cambridge Studies in International Relations, Vol. XVII: Traditions of International Ethics.* New York: Cambridge, 1992. xv, 326 p. Cloth $54.95.

NEWTON-SMITH, W. H., and JIANG TIJANI, eds.: *Popper in China.* New York: Routledge, 1992. viii, 160 p. Cloth $59.50.

NUSSBAUM, MARTHA C., and AMELIE OKSENBERG RORTY, eds.: *Essays on Aristotle's* de Anima. New York: Oxford, 1992. viii, 439 p. Cloth $55.00.

PATTERSON, DENNIS M., ed.: *Wittgenstein and Legal Theory.* Boulder: Westview, 1992. xii, 264 p. Cloth $55.00.

PAUL, ELLEN FRANKEL, FRED D. MILLER, JR., and JEFFREY PAUL, eds.: *Reassessing Civil Rights.* Cambridge: Blackwell, 1991. xi, 223 p. Paper $18.95.

RILEY, PATRICK, ed.: *Essays on Political Philosophy.* Rochester: University Press, 1992. xxviii, 385 p. Cloth $65.00.

ROCKMORE, TOM, and BETH J. SINGER, eds.: *Anti-Foundationalism Old and New.* Philadelphia: Temple, 1992. vii, 251 p. Cloth $44.95.

SCHULTZ, BART, ed.: *Essays on Henry Sidgwick.* New York: Cambridge, 1992. xvi, 421 p. Cloth $65.00.

SHARP, ANN MARGARET, and RONALD F. REED, eds.: *Studies in Philosophy for Children: Harry Stottlemeier's Story.* Philadelphia: Temple, 1992. xvii, 268 p. Cloth $49.95, paper $24.95.

FORTHCOMING PHILOSOPHY TITLES

THE CONSCIENCE OF THE AUTOBIOGRAPHER
Ethical and Religious Dimensions of Autobiography
John Barbour
Studies in Literature and Religion Series
July • 216 pp • ISBN 0-312-08102-2 $59.95

TOWARDS A CRITICAL VICTIMOLOGY
Edited by Ezzat A. Fattah
Aug. • 240 pp. • ISBN 0-312-07551-0 $65.00

POINCARÉ AND THE PHILOSOPHY OF MATHEMATICS
Janet M. Folina
July • 224 pp. • ISBN 0-312-08052-2 $65 00

MORALITY AND SOVEREIGNTY IN THE PHILOSOPHY OF HOBBES
George Shelton
July • 304 pp. • ISBN 0-312-08094-8 $45 00

THOMAS HOBBES
Political Ideas in Historical Context
Johann P. Sommerville
July • 256 pp
ISBN 0-312-07966-4 $45.00 cl.
ISBN 0-312-07967-2 $19.95 pb

THE GREATER GOOD DEFENCE
An Essay on the Rationality of Faith
Melville Stewart
Aug. • 200 pp • ISBN 0-312-07906-0 $39.95

THE THEOLOGY AND PHILOSOPHY OF ELIADE
A Search for the Centre
Carl Olson
July • 225 pp. • ISBN 0-312-07906-0 $39.95

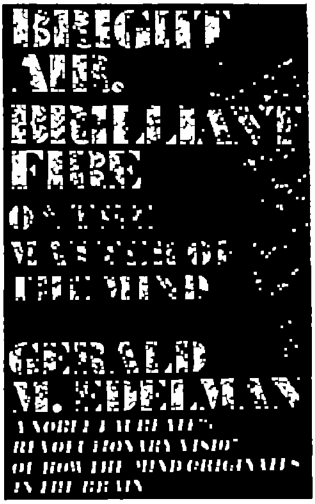
ST. MARTIN'S PRESS
Scholarly & Reference Division, 175 Fifth Avenue, New York, NY 10010

What is the most complex material object in the universe?

Your brain. Now one of the world's foremost brain scientists gives us a glimpse of the revolution in neuroscience that has put us on the brink of understanding ancient mysteries, a revolution that may be as significant as the Galilean and Darwinian revolutions—with implications for philosophy, for curing mental disease, and for building conscious artifacts.

"Convincingly puts together the disconnected pieces of an old puzzle."
—UMBERTO ECO

"Brilliant and daring."
—JEROME BRUNER



$25.00

At bookstores or from
BasicBooks
A Division of HarperCollinsPublishers
Toll-free with credit card 1-800-331-3761
Available in Canada from HarperCollinsCanadaLtd

Human Morality

SAMUEL SCHEFFLER, *University of California at Berkeley*

Some people believe that the demands of morality coincide with the requirements of an enlightened self-interest. Others believe that morality is diametrically opposed to considerations of self-interest. This book argues that there is another position, intermediate between these extremes, which makes better sense of the totality of our moral thought and practice.

1992 160 pp. $26.00

From Morality to Virtue

MICHAEL SLOTE, *University of Maryland*

This book defends a distinctive, intuitive, and symmetric ethical principle according to which we should balance self-concern with concern for others, but it also concludes that there is, contrary to utilitarianism, no single basis for status as a virtue nor any simple relation between the virtues and human well-being.

1992 320 pp. $37.00

Ethics and Excellence

ROBERT C. SOLOMON, *University of Texas, Austin*

Rejecting the myths and metaphors of war-like competition that clouds business thinking and develops an "Aristoteleon" theory of business, the author emphasizes several core concepts: the corporation as community, the search for excellence, the importance of integrity and sound judgment, as well as a more cooperative and humane vision of business.

The Ruffin Series in Business Ethics

1992 304 pp. $29.95

Transforming Vision

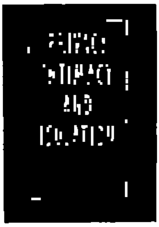
Imagination and Will in Kierkegaardian Faith

M. JAMIE FERREIRA, *University of Virginia, Charlottesville*

"Any serious examination of Kierkegaard will undoubtedly have to respond to her reexamination of the will and how the belief forming faculties operate in their appropriation of ethical and religious reality. This work, along with her earlier *Scepticism and Reasonable Doubt* confirms that Ferreira is one of the most significant thinkers working in the field of philosophical theology and philosophy of religion today."—*The Cresset*

1991 176 pp. $39.95

Privacy, Intimacy, and Isolation



JULIE C. INNESS, *Mount Holyoke College*

By arguing that intimacy is the core of privacy, including privacy law, Inness undermines privacy skepticism, providing a strong theoretical foundation for many of our everyday and legal privacy claims, including the controversial constitutional right to privacy.

1992 176 pp. $24.95

Will and World

A Study in Metaphysics

N.M.L. NATHAN, *University of Liverpool*

Nathan provides a general account of the resolution of the conflict between what we want to be true and what we believe to be true as a philosophical objective, showing that there are ways of thinking it through systematically with a view to resolving or alleviating it and also studies in detail a set of interrelated conflicts about the freedom and the reality of the will.

1992 192 pp. $49.95

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MG
To order by phone using major credit cards please call 1-800-451-7556

Oxford University Press
200 Madison Avenue ✦ New York, NY 10016

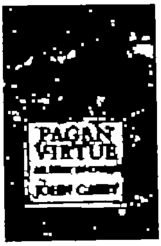
Oxford

New in paperback!

Pagan Virtue

An Essay in Ethics

JOHN CASEY, *Cambridge University*



"Succinct and vigorous....A triumph of common sense and lucidity....Two cheers for Dr. John Casey."—*Sunday Correspondent* "Serves as an excellent layman's guide to some of the most central questions discussed by philosophers....Pagan Virtues deserves some of the highest praise which can be accorded to a work of philosophy: it speaks more directly to the general reader than to the specialist philosopher."—*Financial Times*

1990 (paper 1991) 256 pp. paper $22.95

God's World and the Great Awakening

Limits and Renewals 3

STEPHEN R.L. CLARK, *University of Liverpool*

In the final volume of his triology, Limits and Renewals, Clark defends the primary faith of humankind, that there is a real world which is more than a shadow of our desires and fancies, and which can be discovered through right reason.

1991 256 pp. $55.00

Alfarabi, Avicenna, and Averroes

Their Cosmologies, Theories of Active Intellect and Theories of the Human Intellect

HERBERT A. DAVIDSON, *University of California, Los Angeles*

Davidson examines the cosmologies and theories of human and active intellect in the three philosophers and covers such subjects as: the emanation of the supernal realm from the First Cause; the emanation of the lower world from the transcendent active intellect; prophecy; and human immortality.

1992 384 pp. $39.95

Thought Experiments

ROY SORENSEN, *New York University*

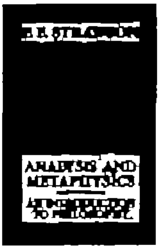
Sorensen presents a general theory of thought experiments: what they are, how they work, what are their virtues and vices. On Sorensen's view, philosophy differs from science in degree, but not in kind. For this reason, he claims, it is possible to understand philosophical thought experiments by concentrating on their resemblance to scientific relatives.

1992 336 pp. $45.00

Analysis and Metaphysics

An Introduction to Philosophy

P.F. STRAWSON, *Oxford University*



In this book, one of the most distinguished of living philosophers explains and illustrates a certain conception of the nature of analytical philosophy. Among the distinctive features of Strawson's exposition are the displacement of an older, reductive conception of philosophical method in favor of elucidating the interconnections between the complex but irreducible notions which form the basic structure of our thinking and the demonstration that the three traditionally distinguished departments of metaphysics, epistemology, and logic are but three aspects of one unified enquiry.

1992 160 pp. paper $12.95 cloth $42.50

Slippery Slope Arguments

DOUGLAS WALTON, *University of Winnipeg*

Walton identifies and analyzes four types of slippery slope arguments and presents guidelines that show how each type can be used correctly or incorrectly, using over fifty case studies of argumentation on controversial issues. These include abortion, medical research on human embryos, euthanasia, and banning of American flag burning.

Clarendon Library of Logic and Philosophy

1992 312 pp. $65.00

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709-M Philosophy Hall, Columbia University
New York, New York 10027

28 OCT 1992

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 9
SEPTEMBER 1992

page
445 *Rationality within Reason* David Schmidtz
467 *Rights between Friends* Michael J. Meyer
484 *Rationality and the Prisoner's Dilemma in David Gauthier's* Morals by Agreement Byeong-Uk Yi

496 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 9, SEPTEMBER 1992

RATIONALITY WITHIN REASON*

Suppose I need to decide whether to go off to fight for a cause in which I deeply believe, or stay home with a family which needs me and which I deeply love. What should I do? My friends say I should determine the possible outcomes of the two proposed courses of action, assign probabilities and numerical utilities to each possibility, multiply through, and then choose whichever alternative has the highest number.

My friends are wrong. Their proposal would be plausible in games of chance where information on probabilities and monetarily denominated utilities is readily available. In the present case, however, I can only guess at the possible outcomes of either course of action. Nor do I know their probabilities, or how to gauge their utilities. The strategy of maximizing expected utility is out of the question, for employing it requires information that I do not have.

Nevertheless, my friends have not given up trying to help, and so they point out that I could simulate the process of maximizing expected utility by assuming a set of possible outcomes, estimating their probabilities, and then making educated guesses about how much utility they would have. I could indeed do this, but I decide not to, for it occurs to me that I have no reason to trust the formula for maximizing expected utility when I have nothing but question marks to plug into it. Better strategies are available. Explaining what they are is the purpose of this paper.[1]

---

* I thank Simon Blackburn, Jim Child, Jules Coleman, Tyler Cowen, Peter Danielson, Walter Glannon, Mark Kingwell, Ruth Marcus, Mark Migotti, Christopher Morris, Jan Narveson, Alan Nelson, Philip Pettit, Ken Presting, Mark Ravizza, Mike Resnik, Lainie and John Ross, Stephen Scott, Michael Slote, Elizabeth Willott, and audiences at Cornell and the University of North Carolina/Chapel Hill for especially helpful comments.

[1] Further questions concern the relationship between rational choice and moral agency. Do we have reasons to be moral? If so, how do they relate to our reasons

0022-362X/92/8909/445–66

I RATIONAL CHOICE

This section distinguishes between optimizing and satisficing strategies, and between moderate and immoderate preferences. The following three sections discuss, in turn, when satisficing strategies are rational, when they are not, and when cultivating moderate preferences is rational. Later sections offer a way of characterizing rational choice in situations where the agent's alternatives are incommensurable.

In the simplest context, one has a set of alternatives clearly ranked in terms of their utility as means to one's ends. If one is an *optimizer,* one chooses an alternative that ranks at least as high as any other. In contrast, if one is a *satisficer,* one settles for any alternative one considers satisfactory. In this static context, though, it is hard to see the point of choosing a suboptimal alternative, even if it is satisfactory.

In a more dynamic and more typical context, we are not presented with a set of nicely ranked alternatives. Instead, we have to look for them, judging their utility as we go. In this context, optimizing involves terminating a search for alternatives upon concluding that one has identified the best available alternative. Although optimizing involves selecting what one judges is best, however, it need not involve judging what is best "all things considered," because sophisticated optimizers recognize that considering all alternatives is not always worth the cost. There may be constraints (temporal, financial, and so on) on how much searching they can afford to do. A person who stops the search upon concluding that prolonging the search is not worth the cost is also employing an optimizing strategy, albeit one of a more subtle variety.[2]

Satisficing, in contrast, involves terminating the search for alter-

---

to be rational? Although I will not address these questions in this paper, they are ultimately the questions which concern me most and which motivate the larger project of which this paper is a part.

[2] Thus, if options emerge serially, one could choose an option in preference to the alternative of searching for something better with no guarantee of ever finding it. The difference between satisficing and this more subtle kind of optimizing has to do with what the two strategies take into account in reaching a stopping point. At any point in the search, we may let the expected utility of stopping the search equal U, the utility of the best option we have turned up so far. The expected utility of continued search equals the probability of finding a better option, $P(fbo)$, multiplied by the utility of finding a better option, $U(fbo)$, minus the cost of further search, $C(fs)$ At some point, the satisficer stops because he believes U is good enough In contrast, the subtle optimizer stops because she believes $P(fbo)U(fbo)-C(fs)$ is less than zero Even if the two stopping rules happen to converge on the same stopping point, they do so for different reasons and require different information

natives upon concluding that one has identified a satisfactory alternative. What distinguishes satisficing from optimizing in the dynamic context is that the two strategies employ different *stopping rules.*[3] Unlike the optimizer, who stops searching when she has either considered all her options or has run up against things like time constraints, the satisficer stops the search upon identifying an alternative as good enough.[4] For example, suppose you enter a cafeteria seeking a nutritionally balanced and reasonably tasty meal. You walk down the cafeteria line surveying the alternatives. If you are satisficing, you take the first meal that you deem nutritionally and aesthetically adequate. If you are optimizing, you continue down the line surveying alternatives until you reach the end of the line or run out of time. You then take the meal that you consider optimal, either in comparison to the other known options or in comparison to the alternative of further search. A satisfactory meal may or may not be optimal. Likewise, as cafeteria patrons know only too well, the best available meal may or may not be satisfactory. Of course, if you switch from one stopping rule to the other, you might end up choosing the same meal, but you will be choosing it for a different reason. So, neither rule is reducible to the other. (One could employ both stopping rules simultaneously, of course, resolving to stop as soon as one finds a satisfactory alternative or runs out of time or has considered all available alternatives—whichever comes first.) Nor can satisficing be equated with the more subtle kind of optimizing that takes the cost of searching for more-than-satisfactory alternatives into account; satisficers select the satisfactory alternative because it is satisfactory, not because they calculate that stopping the search at that point would maximize utility.

---

[3] I borrowed the term 'satisficing' from Michael Slote, *Beyond Optimizing: A Study of Rational Choice* (Cambridge: Harvard, 1989), p. 5, but my characterization of satisficing differs from his. It turns out that my characterization of satisficing as a stopping rule was anticipated by Herbert Simon in *Models of Thought* (New Haven: Yale, 1979), p. 3. In the same passage, Simon also anticipates one of my main complaints about maximization strategies, namely, that they require information that decision makers often do not have

[4] I do not know of any precise way to characterize 'good enough'. Options that promise disease, imprisonment, or premature death are typically held in low esteem, however, so the notion has certain objective elements. But what people consider good enough also seems relative to expectations. As expectations rise, the standards by which an option is judged good enough also tend to rise. This fact can be tragic, for it can rob people of the ability to appreciate how well their lives are going, all things considered. Of course, it is rational to set *goals* with an eye to what is attainable, raising one's sights as higher goals become attainable. But raising the standard by which one deems one's situation satisfactory is harder to fathom. Perhaps people are psychologically incapable of aiming at higher goals without simultaneously reformulating their notions of what is satisfactory. I do not know.

With this characterization of satisficing in mind, we can now clarify the difference between satisficing and *moderation*. Consider that being moderate is contrasted, not with optimizing, but with being immoderate. The distinction between being moderate and being immoderate is a distinction concerning how much one wants. It is a distinction between different ends rather than between different ways of pursuing them. Being a satisficer does not entail that the bundle of goods one would deem satisfactory is of moderate size. A person could be both a satisficer and an immoderate, for a given satisficer may have wildly immoderate ideas about what counts as satisfactory. (Consider a person whose goal in life is to be a millionaire by the age of thirty.) Likewise, that one is an optimizer does not entail that one is immoderate. A person could easily be both a moderate and an optimizer, for the bundle of goods required to maximize her satisfaction may well be of moderate size.

II. WHEN SATISFICING IS RATIONAL

There is an apparent incongruence between the theory and practice of rational choice. Theory models rational choice as optimizing (sometimes even "all things considered") choice, yet in practice satisficing is ubiquitous. We could explain the incongruence away by saying that when people think they are looking for something satisfactory, what they are really looking for is something optimal. But I think satisficing can be reconstructed as a subtle kind of optimizing strategy only on pain of attributing to people calculations they often do not perform (and do not have the information to perform) and intentions they often do not have. This section tries to explain satisficing in terms of thought processes that we can recognize in ourselves. Satisficing will emerge as a real alternative to optimizing, and thus as a strategy that can be evaluated, criticized, and sometimes redeemed as rational.

We begin with the observation that people have a multiplicity of goals. For example, a person can desire to be healthy, to have a successful career, to be a good parent, and so forth. Some goals are broad and others narrow, relatively speaking. While a given pair of goals might have little to do with each other, it is also true that a given goal might be encompassed by another. In other words, the point of the narrower goal—the reason for its being a goal—is that it is part of what one does in pursuit of a broader goal. For example, I might want to upgrade my wardrobe because I care about my appearance because I want a promotion because I care about my career. Suppose I find that the multiplicity of goals that I pursue are ultimately encompassed and rationalized by a goal of making my life as a whole go well. To mark the difference in breadth between my concern for my life as a whole and my concern for particular aspects of my life (such as my health or my career), let us say one seeks a *local* optimum when one seeks to make a certain aspect of one's life go as well as possible. One seeks a *global* optimum when one seeks to make one's life as a whole go as well as possible.[5]

Note that optima can be defined as such only within the context of the constraints under which goals are pursued. (Thus, when economists speak of maximizing utility, it goes without saying that they are talking about maximizing utility subject to a budget constraint.) We pursue goals subject to the limits of our knowledge, time, energy, ability, income, and so on. More intriguing, however, is that we often make decisions under conditions where the set of constraints imposed on us by external circumstances is not determinate enough to yield a well-defined optimum. When this happens, we often impose further constraints on ourselves.

For example, in fleshing out the problem of locating a house, we have to make some prior decisions. We decide how long to look, how much money to spend, what neighborhoods to consider. We knock only on doors of houses displaying "for sale" signs rather than on every door in the neighborhood. To some extent, these constraints are imposed on us by mundane external factors, but they also have a striking normative aspect, for they are in part rules of conduct we impose on ourselves; we take it upon ourselves to make our constraints more precise and more limiting so as to make our choice set more definite. Local optimizing would often be neurotic and even stupid if local goals were not pursued within compartments partly defined by self-imposed constraints. The constraints we impose on our narrower pursuits can keep narrower pursuits from ruining the larger plans of which they are part.[6]

In other words, if we look at life as a whole, we can see that life as a whole will go better if we spend most of it pursuing goals that are narrower than the goal of making life as a whole go better. That is why it is rational to formulate and pursue local goals. But it is also rational to prevent narrower pursuits from consuming more resources than is warranted by the importance (from the global perspective) of achieving those narrower goals. Accordingly, we pursue narrower goals under self-imposed constraints.

---

[5] I borrowed the terms 'local' and 'global' from Jon Elster, although when rereading his text I noticed that the way he uses the terms bears little resemblance to the way they are used here. He says the definitive difference between locally and globally maximizing machines is that the latter, unlike the former, are capable of *waiting* and *indirect* strategies See Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality* (New York: Cambridge, rev. 1984), p. 9

[6] Jules Coleman has pointed out to me that what David Gauthier (*Morals By Agreement* (New York: Oxford, 1986), p. 170) calls constrained maximization is a particularly interesting kind of local optimizing under self-imposed constraints Constrained maximizers seek maximum payoffs in prisoner's dilemmas subject to this constraint: they will cooperate (and thus pass up the opportunity to unilaterally defect) if the expected payoff of cooperating is higher than the known payoff of *mutual* defection, which it will be if and only if they expect their partners to cooperate.

Although the constraints we impose on ourselves are imposed from a more encompassing perspective, it is only within the narrower perspective that we become *subject* to self-imposed constraints. (External constraints, of course, are salient even at the global level.) Self-imposed constraints can be applied only *to* narrower pursuits and can be applied only *from* the perspective of a more encompassing pursuit. Our broader objectives constrain what we are willing to spend on upgrading a wardrobe. They constrain what we will do for the sake of our appearance, or for the sake of a promotion, or for the sake of a career.

Armed with the distinction between local and global optimization, we can now explain when satisficing is rational. Michael Slote believes the optimizing tendency can be self-defeating. He says that "[a] person bent on eking out the most good he can in any given situation will take pains and suffer anxieties that a more casual individual will avoid . . ." (*op. cit.,* p. 40). And he asks us to consider "how much more planful and self-conscious the continual optimizer must be in comparison with the satisficer who does not always aim for the best and who sometimes rejects the best or better for the good enough" (*op. cit.,* p. 40). In short, that one has an opportunity to pursue the good is not by itself a compelling reason to pursue the good. Surely, Slote has an important point. Just as surely, however, his point applies to local optimizing rather than to optimizing as such. From the global perspective, seeking local optima can be a waste of time. Global optimizers seek local optima only when doing so serves their purposes. Thus, satisficing is a big part of a global optimizer's daily routine. Effort can have diminishing returns, so a global optimizer will be careful not to try too hard. Local optimizing often gives way to satisficing for the sake of global optimality.

From the global optimizer's point of view, the process of buying a house provides a good example of how satisficing can be rational. When we choose a house, we might proceed by seeking the best available house within certain constraints—within a one-month time limit, for example. We impose such a limit because we have goals other than living in a nice house. Looking for a house competes with our other goals for our time and energy. Or we might look for a satisfactory house and cease looking when we find one. Most of the people I have asked say they would optimize within constraints, but would not deem satisficing irrational. Like local optimizing, it can serve our larger plans by setting limits on how much effort we put into seeking a house at the expense of other goals that become more important at some point, given the diminishing returns of remaining on the housing market. An optimizing strategy places limits on how much we are willing to invest in seeking alternatives. A satisficing strategy places limits on how much we insist on finding before we quit that search and turn our attention to other matters.

The two strategies need not be inflexible. People sometimes have reason to switch or revise strategies as new information comes in. If I seek a satisfactory house in an unfamiliar neighborhood and am shocked to find one within five minutes, I may stop the search, acknowledging the stopping rule I previously imposed on that activity. On the other hand, I may conclude that, since I formulated my aspiration level under unrealistically pessimistic assumptions, I should resume my search with a satisficing strategy revised to reflect a higher aspiration level. Or I may switch to a local optimizing strategy, spending another day or two looking at houses, and then taking the best I have found so far. Or I may do both, looking until I either reach my new aspiration level or reach my time limit. In this way, the two strategies are often interactive.

Likewise, suppose I started out with the idea of seeking the best house I could find within a one-month time limit but have so far been terribly disappointed with my options. In this case, when after two weeks I finally find a house that meets my plummeting aspiration level, I may find myself embracing that sadder but wiser aspiration level as a stopping rule, abandoning my original plan to seek a local optimum with respect to a one-month time constraint.

I suspect that the more concrete our local goals are, the more reason there is to satisfice. If we do not know exactly what we are looking for, then we usually are better off setting a time limit and then taking what we like best within that limit. But if we know exactly what we are looking for, then it is rational to stop searching as soon as we find something that fits the bill.[7] So, having detailed information about our *goals* weighs in favor of using that information in formulating aspiration levels as stopping rules. Conversely, the more we know about our *set of alternatives,* the easier it is to identify which alternative has the highest utility, which weighs in favor of seeking local optima.

---

[7] Jay Rosenberg tells me that before he began looking for a house, he made a list of desirable features, telling himself he would take the first house that had 85% or more of those features. As it happens, the first house he looked at scored 85% He stopped looking, bought the house, and has lived there ever since

The stakes involved are also pertinent—indeed crucial. The less we care about the gap between satisfactory and optimal toothpaste, for example, the more reason we have to satisfice—to look for a satisfactory brand and stop searching when we find it. Note the alternative: instead of satisficing, we could optimize by searching among different brands of toothpaste until we find the precise point at which further search is not worth its cost. But an optimal stopping point is itself something for which we would have to search, and locating it might require information (about the probability of finding a better brand of toothpaste, for example) that is not worth gathering, given the stakes involved in the original search for toothpaste. Against this, one might object to my assumption that we need precision in the search for an optimal stopping point. Why not seek to learn *roughly* when looking for better toothpaste is not worth the cost? In the search for a stopping point that we might graft onto the original search for toothpaste, it can be more rational to seek to be tolerably close to an optimum than to seek to be at an optimum.

But that is my point: there are cases where one does not care enough about the gap between the satisfactory and the optimal to make it rational to search for the optimal. Searching for optimal toothpaste can be a waste of time, but so can searching for the optimal moment to quit looking for toothpaste. One way or another, satisficing enters the picture. There will be times when even the most sophisticated optimizing strategies will be inappropriate, for they require information that we may not have and whose acquisition may not be worth the trouble. And a less sophisticated "all things considered" strategy will nearly always be inappropriate. Rational choice involves considering only those things which seem worthy of consideration, which is to say it involves satisficing, i.e., having a stopping rule that limits how comprehensive a body of information we insist on gathering before stopping the search and turning our attention to other matters.

There is also something to be said for having a moderate disposition—a disposition that allows one to be content with merely satisfactory states of affairs. Consider that starting a search too soon can be every bit as wasteful as stopping a search too late. Searching for a house is costly. It is costly partly because people have other goals; the time and energy you spend searching for a house could have been spent on other things. Even if you find a better house than you already have, the process of moving will also be costly. Moreover, it takes time living in and enjoying a house in order to recoup these costs. If you move every month, you will always be paying the costs and never enjoying the benefits of better housing. Moving into a house is part and parcel of a decision to stay a while, for it is only in staying that you collect on the investment of time and energy you made in moving. The general lesson is that costly transitions to preferred states of affairs require intervening periods of stability so that transition costs can be recovered and thus rationally justified. The stability of the intervening periods requires a disposition to be content for a while with what one has—to find something one likes and then stop searching.

Further, even if transition costs are relatively minor, there still can come a point when we should abandon the search for, say, a better job or a better spouse, not because such goals are unattainable or even because the transition costs are too high, but rather because such goals eventually can become inappropriate. At some point, we have to start collecting the rewards that come only when we make a genuine commitment—when we stop looking for something or someone better. Local optimization processes have their own ends, but initially adopting an end and creating a compartment within which to pursue it is itself a goal-directed activity and, from the standpoint of the global optimizer, not one to be engaged in frivolously. We need to be able to satisfice within various local compartments (those defining our searches for spouses, jobs, and so on) in order to be able to make our lives as a whole go as well as possible.

III. WHEN SATISFICING IS NOT RATIONAL

Slote says that "choosing what is best for oneself may well be neither a necessary nor a sufficient condition of acting rationally, even in situations where only the agent's good is at stake" (*op. cit.*, p. 1). For example, a person who is moving and must sell his house might seek, "not to maximize his profit on the house, not to get the best price for it he is likely to receive within some appropriate time period, but simply to obtain what he takes to be a good or satisfactory price" (*op. cit.*, p. 9). Once the seller receives a suitable offer, he may rationally accept it immediately, even though there would be no cost or risk in waiting a few days to see if a higher offer materializes. "His early agreement may not be due to undue anxiety about the firmness of the buyer's offer, or to a feeling that monetary transactions are unpleasant and to be got over as quickly as possible. He may simply be satisficing in the strong sense of the term. He may be moderate or modest in what he wants or needs" (*op. cit.*, p. 18).

Slote does not offer an analysis of rationality. Nor shall I. I do, however, offer this as a necessary condition of rationality: one's choice is rational only if one did not recognize clearly better reasons for choosing any of one's foregone alternatives. This condition does not beg any of the questions that concern us here. It does not entail that rational choice is optimizing choice. Rather, it allows that one could rationally choose an alternative because it is satisfactory, terminating the search of one's choice set at that point.[8] Moreover, it also allows that, if one has two satisfactory alternatives, one could choose the more moderate of the two on the grounds that it satisfies a preference one happens to have for moderation.

On the other hand, although a suboptimal option may be good enough to be worthy of choice in a given case, that does not mean it is worthy of being chosen in preference to something better. If one has two choices and one alternative is satisfactory but the other is not, then the satisfactory choice is rational because it is *better*. But suppose one has two choices and both are satisfactory. (E.g., suppose your house is for sale, and you simultaneously get two satisfactory offers, one for $200,000 and another for $210,000, and you prefer the larger offer.) In this case, one cannot rationalize choosing the inferior alternative merely by pointing out that the inferior alternative is satisfactory. The inferior option is satisfactory, but since this is not a difference, it cannot *make* a difference either. By hypothesis, the superior option is also satisfactory.

Why, then, should we choose the superior option? Presumably because it is better. Whatever it is in virtue of which we deem that option superior is also a reason for us to choose it.[9] Oddly, Slote denies this. It can be rational to choose the inferior option, Slote insists. Nor do we need a reason to choose the inferior option, Slote argues, because rationality does not always require people to have a reason for choosing one alternative rather than another (*op. cit.*, p. 21). For example, a rational person might grab a blouse out of her closet in the morning without being able to explain why she chose that one over the similar blouses hanging beside it. To call her irrational simply because she cannot explain her choice would be a mistake.

---

[8] That is, an optimizer might choose a satisfactory option in preference to searching for better options that might never materialize. Slote, however, says we intuitively recognize the rationality of taking the first satisfactory offer even in abstraction from the real-world risks and anxieties of having to sell one's house (*op. cit*, p 18). But my view is that, if we are going to talk about common sense allowing the seller to accept the firm offer immediately even though the seller has the option of waiting a few days in hope of a higher offer, then we have to stick to the conditions under which common sense holds sway. We do indeed have intuitions about what we should do in risky situations, but we cannot, as Slote wants to do, simply *stipulate* that our intuitions regarding risky situations have nothing to do with the fact that in the real world such situations are risky In the real-world housing market, to turn down an entirely satisfactory offer in quest of something better is to court disaster, to tempt fate, as it were This is one reason why it is common sense, and rationally explicable common sense, for a global optimizer to be hesitant about turning down a satisfactory offer. Even from a local perspective, the expected gain from further search may not be worth risking the potential loss.

[9] Philip Pettit makes the same point in "Satisficing Consequentialism," *Proceedings of the Aristotelian Society*, supplement to LVIII (1984): 165–76, here p. 172

I agree with this, as far as it goes; not all choices have to be or can be explained. To deem a choice rational, however, is to imply there is an explanation of a certain kind. A person can be rational without being aware of reasons for everything she does, but the things she does for no reason are not rational, and we do not show them to be rational merely by pointing out that they were done by a rational *person*. The person who simply grabs a blouse may be choosing, perhaps rationally, to forgo the opportunity to choose rationally which of her several blouses she ends up wearing. If she is running late for the train, then under the circumstances anything that counts as a blouse will also count as satisfactory, so she leaves to impulse the selection from her set of blouses. (In this case, the process of searching among alternative blouses virtually vanishes—there is hardly any choice at all. If she instead gives herself a few seconds to make sure she avoids blouses with valentines or political slogans on them, then she will be choosing within a small but real local compartment.)

Thus, it can be rational to forgo the opportunity to choose rationally, because searching for reasons to choose one alternative over another might be a waste of time from a broader perspective. Nevertheless, it definitely is a condition of rational *choice* that the choice in question is made for a reason. Deliberately choosing the worse over the better is irrational, and this is not mitigated by the fact that rational people sometimes leave their choices to impulse.

Rational-choice theory can tell us a story about why the commuter finds herself going to work in a green blouse with orange polka dots, but the story will require an implicit or explicit distinction between more and less encompassing perspectives. Without the distinction, an optimization story would be blatantly false, for she does not in fact choose the optimal blouse, and a satisficing story would have neither explanatory nor justificatory power, for the point of choosing a merely satisfactory blouse when better ones were available would remain a mystery. To be able to see the point of what she does at the local level, we have to be able to step back and look at her actions from a broader perspective. From a broader perspective, it may be perfectly reasonable for her to cut short a particular search because she was looking for something satisfactory and found it, but it cannot be rational to choose something because it is satisfactory while at the same time recognizing a clearly better alternative.

IV. WHEN MODERATION IS RATIONAL

As I said earlier, we cannot choose between two satisfactory alternatives on the grounds that one of them is satisfactory. We could, however, choose on the grounds that one of them is more moderate. Consider an example of Slote's. He says it "makes sense" for someone to desire "to be a really fine lawyer like her mother, but not desire to be as good a lawyer as she can possibly be. This limitation of ambition or aspiration may not stem from a belief that too much devotion to the law would damage other, more important parts of one's life. In certain moderate individuals, there are limits to aspiration and desire that cannot be explained in optimizing terms . . ." (*op. cit.*, p. 2).

Common sense indeed recognizes that moderate aspirations can be rational, but to note this fact in an off-the-cuff way is hardly to provide an explanation of moderate aspirations. Our common-sense recognition is precisely what has to be explained. If all we have is an intuition that an act makes sense, but cannot say what the act makes sense *in terms of*, then we would be jumping to conclusions if we said we were approving of the act as rational. In contrast, if we explain a show of moderation in terms of its conduciveness to overall satisfaction, then we have explained it as rational. We have not merely claimed that it makes sense; rather, we have actually made sense of it. We have shown that we had reason to choose as we did, while not having better reasons to choose differently.

How, then, might we explain having moderate career goals? First, there is the issue of trade-offs mentioned by Slote. One might cultivate an ability to be content with moderate career goals, not because one prefers moderate success to great success, but because one cares about things other than success. Thus, one point of cultivating modest desires with respect to wealth is that it might improve a person's ability to adhere to a satisficing strategy with respect to income, thus freeing herself to devote time to her children, her health, and so on.

There are also ways in which moderation can have instrumental value that do not depend on the need to make trade-offs. There can be reasons for striving to be as good a lawyer as one's mother even if one wants to be as good a lawyer as possible. For example, a person might aim at being as good as her mother as a stepping stone to *becoming* the best lawyer she can be. The modesty that enables a person to concentrate on successfully making smaller steps may eventually put her within reach of something more lofty. There is also value in concreteness. A person may have no idea how to go about becoming the best possible lawyer, but may have a much clearer idea about how to become as good as her mother because the more modest goal is more concrete. Further, even given two equally concrete goals, an optimizer might very well choose the lesser on the grounds that only the lesser goal is realistic. Thus, one might become a better lawyer by emulating one's highly competent mother than by wasting one's time in a fruitless attempt to emulate her superstar partner.

Finally, we can at least conceive of moderation being a preference in itself—not just a quality of a desire but itself the thing desired.[10] One might explain the cultivation of such a preference on the grounds that moderation is less distracting than extravagance, with the consequence that the moderate life is the more satisfyingly thoughtful and introspective life. In various ways, then, moderation can have instrumental or even constitutive value from the global perspective. Insofar as moderate preferences can be deliberately cultivated, their cultivation is subject to rational critique, and can thus be defended as rational.

V GLOBAL PERSPECTIVE AND UNDERDETERMINED RATIONAL CHOICE

Like moderate preferences, satisficing strategies can be of instrumental value from the global perspective. Satisficing strategies, however, can be of instrumental value only. This is because to satisfice is to give up the possibility of attaining a preferable outcome, and giving this up has to be explained in terms of the strategic reasons one has for giving it up. Local optimizing must likewise be explained, for it too consists of giving something up, namely, the opportunity to invest one's efforts in some other compartment.

Global optimizing, however, is not open to question and subject to trade-offs in the ways that local optimizing and satisficing are. Local goals can compete with each other, but there are no goals that compete with optimizing at the global level, at least not in the arena of rationality. A global optimum is not one among several competing goals; rather, in encompassing our lives as a whole, it also encompasses our competing goals. It represents the best way to resolve the competition from the standpoint of life as a whole. Local optimizing can be a waste of time from the global perspective, but global optimizing cannot.

What, then, is the nature of the global perspective? Do we ever actually assume the global viewpoint or is this merely a theoretical postulate? The answer is that we can and do assume the global viewpoint every time we do what we call "stepping back to look at the big picture." We do sometimes ask ourselves if the things we do to advance our careers, for example, are really worth doing. We do not spend all our waking hours looking at the big picture, of course. Nor should we, for when we look at the big picture, one thing we see is that we can spend too much time looking at the big picture. Reflection is a crucial part of the good life, but it is only a part. Part of attaining a global optimum involves being able to lose ourselves for a time in our local pursuits.

---

[10] I thank Mark Ravizza for this point.

To speak of attaining a global optimum, however, and to endorse global optimizing as a normative ideal, is not to presume that there is an algorithm for identifying global optima. In fact, not every set of alternatives even contains a well-defined optimal choice, let alone one that can be easily identified. To borrow a fanciful example from John Pollock,[11] suppose you are immortal, and are also fortunate to have in your possession a bottle of "EverBetter Wine." This wine improves with age. In fact, it improves so steadily and so rapidly that no matter how long you wait before drinking it, you would be better off, all things considered, waiting one more day. The question is: When should you drink the wine?

A rational *person* presumably would simply drink the wine at some point (perhaps after artificially constraining himself to drink the wine by year's end, and then picking New Year's Eve as the obvious choice within that time frame), but the person would not be able to defend any particular day as an optimal choice. Indeed, it is built into the example that no matter what day the immortal chooses, waiting one more day would have been better. There are no constraints with respect to which he can regard any particular day as the optimal choice, unless he imposes those constraints on himself.

There is something about choosing New Year's Eve that is rational, but it consists of something other than how that day compares to the alternatives. Although the immortal could not defend choosing New Year's Eve in preference to waiting one more day, the choice is defensible in the sense that he did not have a better alternative to picking *something or other*. Indeed, picking something or other was the optimal selection from the set of options whose only other member consisted of sitting on the fence forever. The distinction between local and global optimizing thus allows us to explain without paradox the sense in which choosing New Year's Eve was in our lives only when they develop a certain autonomy, when we pursue them not as means of making our lives go well but as ends in themselves. We begin to tap the capacity of our ideals, our spouses, and our children to enrich our lives only when we acknowledge them as having value far beyond their capacity to enrich our lives. (Cherishing them becomes more than an instrumental means of making life go well; it becomes constitutive of life going well.) And goals that we come to cherish as ends in themselves inherently tend to become *incommensurable with each other*. We may, for instance, find ourselves in a position where we cannot fight for a cause in which we deeply believe without compromising the care which our loved ones need from us and which we wholeheartedly want them to have. Nevertheless, this is the price of the richness and complexity of a life well-lived. To have both ideals and loved ones is to run the risk of having to make decisions under unresolved conflicts of value.

---

[11] "How Do You Maximize Expectation Value?" *Noûs*, XVII (1984): 409–21, here pp. 417ff.

Because some of our values are incommensurable, we sometimes have no method by which to identify optimal trade-offs among conflicting local goals. In such cases, the goal of making life as a whole go as well as possible remains meaningful, although there may not be any course of action that unequivocally counts as pursuing it. Even if would-be global optimizers cannot identify optimal options, they can still reject alternatives that fail to further any of their goals. In particular, if no better way of resolving the conflict emerges, simply picking something or other will emerge as optimal compared to the alternative of remaining on the fence, for we eventually reject fence sitting on the grounds that it fails to further any of our goals.

This may seem a grim picture of rational choice at the global level, but there are two points to keep in mind. First, when faced with a situation in which we must simply pick something, we are likely to have regrets about paths not taken, but we naturally adapt to the paths we take, and regret can fade as we grow into our choice. Thus, an alternative somewhat arbitrarily picked from a set within which no optimum exists can eventually come to be viewed as optimal from the perspectives of people we are yet to become, even if it could not be considered optimal at the moment of choice. Second, this discussion of underdetermined rational choice concerns a worst-case scenario. Global optimizers carry out the highest-ranked life plan when they have one. Often, however, there is no highest-ranked plan for life as a whole and thus no well-defined global optimum; there is only a need to cope with competing and sometimes incommensurable local goals. In the worst case, no course of action unambiguously qualifies as making life as a whole go as well as possible, except insofar as it is unambiguously better to move in some direction rather than none. But this gives us enough to avoid paralysis even in the worst case. By hypothesis, simply picking something emerges as the best the agent can do, and thus to pick something is to optimize with respect to the choice of whether to spend more time sitting on the fence.

It would be natural to say rational choice is choice "all things considered." The trouble is that we often find ourselves not knowing what to consider, and it would be bad advice to tell us to consider all things. We can consider all things within a limited range, perhaps, but the limits of that range will themselves tend to be matters of choice in large part. We start out knowing that in some sense we want each aspect of our lives to go as well as possible, yet we realize that our resources are limited and that our various pursuits must make room for each other. When looking at our lives as a whole, what is most clear is that rationally managing a whole life involves managing trade-offs among life's various activities. If the benefits that will accrue from our various pursuits are known and commensurable, then managing the trade-offs is easy, at least theoretically; we simply maximize the sum of net benefits. In many of the everyday cases I have been discussing, however, the benefits are neither known nor commensurable with other benefits. Even so, we can effectively manage trade-offs among our particular pursuits by setting limits on how much of our lives we will spend on particular pursuits. We can also set limits on how much benefit we insist on getting from particular pursuits. To impose the latter kind of stopping rule on a particular pursuit is to embrace what I have called a satisficing strategy.

Both kinds of constraint play a role in rational choice. Why? Because if we recognized only temporal limits, say, then we would automatically spend our full allotment of time in a given compartment even when we had already found an acceptable option. But if we also have strategically limited aspiration within that compartment, then finding an acceptable option will trigger a second kind of stopping rule. The second stopping rule closes the compartment and diverts the unused portion of the compartment's time allotment to other compartments where our need to find an acceptable option has not yet been met. Cultivating moderate preferences may also be advantageous in a supplementary way insofar as moderate preferences may help us adhere to the kind of limit we impose on a pursuit when we embrace a satisficing strategy.

Against the idea that our most important goals tend to become incommensurable with each other, one might suppose our global end is simply to flourish or to be happy—and that our local goals therefore *must* be commensurable in such terms. This would be a tidy climax to an otherwise rather untidy story about rational choice under unresolved conflict of values, but the tidiness would be superficial. One hardly gives people an algorithm for resolving conflicts when one advises them to be happy. What makes such advice vacuous, I think, is that flourishing and being happy cannot be concrete goals at the global level in the way that finding a house can be at the local level. Of course we *want* to flourish, but we *aim* to flourish only in an especially metaphorical sense. The fact is that we flourish not by aiming at flourishing but by successfully pursuing things other than flourishing, things worth pursuing for their own sake.

Likewise, happiness can be a standard by which a life as a whole is judged, perhaps, but it cannot be a goal at which a life as a whole is aimed. We do not become happy by pursuing things there would otherwise be no point in pursuing. Rather, there must be a point in striving for a certain goal before striving for it can come to have any potential to make us happy. To aim at happiness is to aim at a property that can emerge only in the course of aiming at something else.[15] So, the point about happiness and flourishing leaves us where we started, having to choose among things we value for their own sake, hoping we will be happy with our choice.

I might add that happiness derives from a variety of local sources, and the different elements of a person's happiness are not interchangeable. We can find happiness in our careers or in our marriages, but the vacuum left by a shattered career cannot be filled by domestic bliss. Our various local pleasures are not fungible. They do not admit of an overall measure suitable for use at the global level.[16]

VI. THE BROADEST PERSPECTIVE

The global perspective is the perspective that encompasses our lives as a whole. Decision making at this level disciplines the amount of time we devote to particular local compartments. I think we are capable of taking a perspective this broad even in worst-case scenarios where there is no well-defined global optimum. But even if we suppose that we can take a perspective encompassing our whole lives, why should we suppose that this is the broadest perspective we can take?

---

[15] As Bernard Williams puts it, one has to want other things for there to be anywhere that happiness can come from. See "A Critique of Utilitarianism," in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (New York: Cambridge, 1973), p. 113.

[16] I thank Nick Sturgeon for a discussion from which this point emerged

Perhaps there can be broader perspectives than what I have called the global perspective. Indeed, in the larger project of which this paper is a part, I hope to show that we do have access to a larger perspective, that there are some aspects of morality that we cannot appreciate except from this larger perspective, and that it can be rational to try to achieve this perspective. On the other hand, I think it would be highly unrealistic to suppose there is an infinite regress of levels. There is no need to prove that an infinite regress is impossible, but I would like to be able to show that my theory does not *presuppose* an infinite regress.

The threat of infinite regress arises in the following way. I said we cannot spend all our time looking at life as a whole; we must be able to lose ourselves (or perhaps I should say, find ourselves) in our local pursuits. How much time, then, should we spend pondering conflicting values? How much time should we spend looking at life as a whole? From what perspective do we decide what constraints to apply to the amount of time we spend looking at our lives from the global perspective? Perhaps we need a "super-global" perspective in order to answer these questions. After all, how could we decide how much time to spend at a given level unless we did so from a still more encompassing perspective? It seems that my theory can explain the time we allot to a given perspective only by supposing that we retreat to a broader one, ad infinitum.

But the theory presumes no such retreat. There are simpler, more realistic ways to explain the amount of time we spend looking at life as a whole.[17] First, there are things, like sleeping, that we do as the need arises; since we do not *decide* how much time to spend sleeping, we do not decide from a broader perspective, either. Indeed, we might be better off sleeping as we feel the need rather than trying to set aside a calculated amount of time for sleep. Perhaps the same holds true of the activity of looking at life as a whole. Insofar as our purpose in looking at life as a whole is to resolve conflicts arising between various aspects of our lives, so that life as a whole may go as well as possible, there will come a time when taking a global perspective has served its purpose. At that time, the compartment reserved for the activity of resolving local conflicts naturally closes until subsequent conflict forces it open again. There is no residual conflict awaiting resolution at a higher level. Thus, the question of how much time to spend in contemplation need not itself require contemplation. Rather, we take whatever time it takes to resolve a conflict, or else we reach a point where we must simply pick something. More generally, we stop contemplating when we judge that pursuing our local goals has come into conflict with—and has become more important than—the activity of thinking about how to juggle them. (For example, we would not dwell on the big picture if we were starving. Conflicts are rarely so important that contemplating them could pre-empt securing our immediate survival.) In this scenario, we are driven *to* the global level by local conflict and eventually are driven *from* that level by a need to get on with our lives.[18]

---

[17] The simplest way to explain the amount of time spent at the global level would be to say that we take whatever time we need to consider *everything*. The trouble is that we do not have time to consider everything that might be relevant to life as a whole, any more than we have time to consider everything that might be relevant to the purchasing of a house. The explanation will have to be more complicated than this, hence the line of thought pursued in the following text.

But we can also imagine a second kind of scenario in which the amount of time we spend looking at life as a whole is determined by deliberate calculation, in the same way that we could imagine deliberately calculating how much time to spend sleeping. Could we make a conscious decision of this kind without taking a super-global perspective? Yes, we could. Consider that contemplation is an activity that must find its place in our lives along with other activities. For example, I might spend the month of June in a rented cabin, not doing anything to pursue my career, but just thinking about why I ever wanted to be a philosopher and about whether my original reasons still hold. This compartment in my life is reserved for contemplating my career. It is separate from the compartment or compartments within which I actually pursue my career. I also have a compartment, similar in many respects, within which I contemplate life as a whole. But although the *subject* I contemplate is the whole of my life, the contemplation itself is not. The contemplation is only one of many activities about which I care.

Now, if I need to decide how much time to reserve for contemplating life as a whole, I take a global perspective, trying to gauge how important that activity is to my life as a whole. Notice, then, what is unique about the compartment I reserve for the activity of contemplating my life as a whole. The compartment is unique because its boundaries are set by the activity that takes place within it. In the course of contemplating life from the global perspective, I decide how much time to reserve for any given activity, including contemplation in general and contemplating life as a whole in particular. In this scenario, like the previous one, no boundary-setting issue is left to await resolution at a higher level.

---

[18] We also can be driven to a global perspective by the resolution of conflicts Thus, when we finish a major project that had forced other pursuits to take a back seat, we often take time to evaluate self-imposed constraints and decide how to divide our extra time among previously neglected projects. And what drives us from the global perspective is the eventual resolution of a local conflict between savoring the big picture (a satisfying activity indeed when just finishing a major project) and the need, say, to start making dinner

I have outlined two possibilities here. In one case, I use whatever time it takes to resolve conflicts, subject to pre-emption by activities that in the short run are more important than conflict resolution. In this case, no decision is required. The discipline is automatic. In the second case, I discipline the compartment from within, as the process of contemplating trade-offs culminates in time being reserved for contemplation along with other local activities. Therefore, I do not need a super-global perspective to decide how much time to reserve for the activity of taking the global perspective. Such decisions are precisely the kind I make from the global perspective itself, if I need to make them at all. Unless we introduce something that competes with the goal of making life as a whole go as well as possible (such as, perhaps, the recognition of moral obligations), there is no reason to step back from a global perspective to something even broader.

VII. CONCLUSIONS

This paper sets out a normative ideal of rational choice that is suitable for the kind of beings we happen to be, beings who would only hurt ourselves if we tried to maximize our overall utility in every waking activity. It defines satisficing and local optimizing as strategies for pursuing goals within constraints that are in part self-imposed. Satisficing emerges not as an alternative to optimizing as a model of rationality, but rather as an alternative to local optimizing as a strategy for pursuing global optima.

Under normal conditions, we employ a combination of heuristics, such as (1) compartmentalizing our pursuits so as to narrow the scope of any particular optimization problem to the point where our limited knowledge becomes sufficient to identify an optimal solution, (2) accepting self-imposed constraints for the same reason as well as to keep particular pursuits from pre-empting more important ones, and (3) satisficing, which has the effect of closing compartments as soon as they serve the purpose for which they were created. Under normal conditions, where we lack the information we need to assign probabilities and utilities, this combination of strategies is more effective at making our lives as a whole go well than the alternative of plugging guesswork into a formula for maximizing expected utility. Thus, it is no wonder we so rarely make any attempt to calculate expected utilities, for the truth is that we usually have better things to do.

When goals are in conflict, there may not be any well-defined sense in which one way of resolving the conflict is better than the alternatives from the viewpoint of one's life as a whole. Of course, we do well to cultivate moderate preferences so as to reduce the frequency and severity of conflicts of value. But at the same time, there are limits to what one should do to avoid situations of underdetermined choice, for the risk of finding oneself in such situations is a risk we assume in the process of becoming rationally committed to particular ends as ends in themselves. A life without regrets (i.e., regrets about decisions made under unresolved conflict) is preferable all other things equal, but if the lack of regret is purchased at a cost of not having goals that can come into unresolvable conflict, the price is too high. To adopt a number of goals as ends in themselves is to risk finding oneself in situations where global optima do not exist, but there are reasons why a global optimizer would take that risk.

I admit that the theory is not particularly neat and tidy, certainly not in comparison to the standard model. But tidying up the theory at the expense of realism would be a mistake, for the theory is meant to be about us, not about mathematically tractable caricatures of us. I believe rational-choice theory developed along the lines indicated here has more power than standard maximization models to explain the ways in which we actually live, but it does not thereby become merely a self-congratulating description of how we live. Rather, it remains (or becomes) a tool for evaluating and criticizing the ways in which we actually live. It sets out a normative ideal of rational choice that it would be natural and healthy for us to try to live up to.

DAVID SCHMIDTZ

Yale University

RIGHTS BETWEEN FRIENDS*

Are there rights between friends? The very question will sound odd to some, even perverse to others. Some might insist that friendship is one of the few human relationships that can be carried on successfully without any recourse to rights. People only have rights against those who are indifferent, or actively antagonistic, to their legitimate claims. Between true friends rights are simply useless, or possibly their use is even offensive. Even more strongly, others might contend that friendships should be carried on without the admixture of the defensive, aggressive, and individualistic notion of rights. Indeed, perhaps introducing the idea of rights (or allied concepts like claims, obligations, or protection) into an account of friendship will inevitably undermine the very notion of friendship. Along this line, some might also assert that in practice the self-conscious possession of rights will surely lead to the demise of any particular friendship.

This paper will address these questions: Are there rights between friends or does the adversarial nature of rights inevitably frustrate the truly mutual orientation of friendship? If the case can be made that there are rights between friends, just what kinds of claims might friends make against one another? To assume simply that there is no possible legitimate employment of rights between friends is to entertain some common, and I shall argue misguided, presumptions about both rights and friends. For instance, one must not regard it as a given that the existence of a friendship directly results in a flawless cooperative partnership—a kind of "preestablished harmony" between friends. Equally, one must not presuppose that the activity of claiming rights is not liable to any limitations. Such presumptions lead to a less than sophisticated view of such rich and complex notions as rights and friendship.

The central question here—Are there rights between friends?—can also be seen as one form of a slightly broader question: Are there rights within close-knit communities? Part of an increasingly popular "communitarian" critique of a rights-based, "liberal" tradition in political theory is that in certain kinds of communities—

---

* A version of this paper was read at the Seventh Annual Meeting of the North American Society for Social Philosophy and the 1992 Pacific Division Meeting of the American Philosophical Association; I would like to thank discussants there for comments. I also thank Neera K. Badhwar, Marcia Baron, Don Beggs, Martin Cook, Jim Felt, Martin Gunderson, Philip Kain, Elizabeth Radcliffe, and Dan Radcliffe, for helpful comments on earlier drafts of this paper.

0022-362X/92/8909/467–83

especially close-knit ones like friendships where shared purposes and mutual togetherness are the rule not the exception—rights (and kindred notions like justice) are not at all valuable.[1] Indeed, having rights may even be harmful to such communities because they inevitably promote noncommunitarian values like egoism and individualism, and perpetuate antagonistic attitudes where they have no place. I do not plan to address directly the somewhat larger question—Are there rights in *all* types of communities? But given the modest assumption that friendship is a type of community (explained and defended later), some gain in thinking about political theory may follow, if we conclude that it is sensible to think that persons have rights within the close-knit community that arises between friends. At least it may in this way be easier to see how having, and at times exercising, rights is not antithetical to the values and virtues of even the very closest of communities.

I

It seems useful to start by stipulating some general working definitions of the central concepts in question. For the purposes of the following discussion: (1) "Tad has a right" if and only if "Tad has a claim protected by (or correlated with) an obligation"; (2) "Tad has a right against Tim" if and only if "Tad has a claim protected by Tim's obligation to honor Tad's claim." This characterization of rights is central to one major tradition of understanding rights.[2] It is worth noting one further quite simple distinction here: Tad may have a legal right or a moral right against Tim. Tad's *legal* claim, based say on a valid contract, would be protected by Tim's *legal* obligation—for example, to deliver some commodity to Tad on a specified date. Tad's *moral* claim would, in turn, be protected by Tim's *moral* obligation—for example, not to use Tad's private diary for his summer reading at the beach. To have a right against someone in this general sense is simply to have a legitimate claim (either legal or moral or both) protected by their (legal or moral) obligation(s). For the present, other complexities often deemed relevant to the analysis of rights talk (for example, powers, privileges, liberties) can be set aside.

---

[1] See for example, Michael J. Sandel, *Liberalism and the Limits of Justice* (New York: Cambridge, 1982), esp. pp. 28–35; Carol Gilligan, *In a Different Voice* (Cambridge: Harvard, 1982), esp. pp. 100, 136f, 149, 164f, 174; John Hardwig, "Should Women Think in Terms of Rights?" *Ethics*, XCIV, 3 (April 1984): 441–55; Robert Young, "Dispensing with Moral Rights," *Political Theory*, VI (1978), esp. pp. 67–8; Karl Marx, "Critique of the Gotha Program," in *The Marx Engels Reader*, R. Tucker, ed. (New York: Norton, 1978), pp. 525–41. For a good discussion of the view that Marx's view of the ideal community, the highest stage of communism, would forego or transcend rights, see Philip J. Kain, *Marx and Ethics* (New York: Oxford, 1988), pp. 12, 186, 200, 202f.

[2] See, for example, Joel Feinberg, "The Nature and Value of Rights," *Journal of Value Inquiry*, IV (1970): 243–57; A. I. Melden, *Rights and Persons* (New York: Oxford, 1977), pp. 3ff; John Stuart Mill, *Utilitarianism* (Indianapolis: Hackett, 1979), p. 52; W. D. Ross, *The Right and the Good* (New York: Oxford, 1930), p. 50, Judith J. Thomson, *The Realm of Rights* (Cambridge: Harvard, 1990), esp. pp. 37–60; Richard Wasserstrom, "Rights, Human Rights and Racial Discrimination," this JOURNAL, LXI, 20 (October 29, 1964) 628–41.

In this paper I shall talk of a claim *protected* by an obligation rather than one simply *correlated* with an obligation. I do so principally because the notion of protection helps emphasize what is often regarded as the defensive posture represented by the concept of rights. The notion of "protection" is meant to do the same logical job as "correlation"—that of directly linking Tim's claim to Tad's obligation. This characterization, however, does suggest something about the nature of that link. Tad's obligation is not merely joined indifferently (that is, simply correlated) to Tim's claim; it offers some promise of protection in the case of a transgression. The practice of claiming rights is clearly related to the protection of (or the justified attempt to protect) one's interests. This use of the notion of protection is not put forth here as an independent argument about the idea of rights. Those who prefer 'correlate' to 'protect' can make the substitution without disturbing the present argument. I simply hope to use an account of 'having a right' which is coherent as well as sufficiently strong enough to grant credence to the initial problem raised by many of the communitarian critics of rights.[3] The goal at this point in the argument is simply to avoid begging the question in favor of the conclusion—roughly, that friendship and rights are fully congenial ideals—by watering down the notion of rights (or friendship) so as to make the case for their compatibility essentially a foregone conclusion.

Friendship is also a complex social and moral phenomenon. For the purposes of this paper the following characterization, roughly Aristotelian in inspiration, will be helpful to get the discussion started.[4] Jill and Jane are each other's friends when both Jill's and Jane's own well-being has as one central constitutive feature the well-being of the other. As friends, both Jane and Jill have as one of their ongoing concerns the welfare and happiness of the other, and this concern for the other is, in large part, for the sake of the other. In the words of Lawrence Blum's[5] recent account of friendship, friends have "a disposition to act to foster [the good of the friend] simply because the other is one's friend" (*ibid.*, p. 43). Friends at least share a disposition to engender those conditions which provide for the flourishing of both the friend as well as the friendship. Like Aristotle's, this characterization of friendship has the mark of an ideal. So, it is best to note at the outset that it might not always (or often) be fully realized in what we commonly call friendships. We are discussing here what might better be called a true or a genuine friendship.

---

[3] The notion of rights as claims is also important here because the centrality, and complexity, of the *activity* of claiming rights calls for close investigation. Also, one must not confuse protecting the *claim* and protecting the *friendship*—the former may not always result in the latter. Claiming rights may only have a remedial effect on the friendship, or it may, but it need not, diminish it. See sections III and IV below.

[4] *Nicomachean Ethics*, Terence Irwin, trans. (Indianapolis. Hackett, 1985), bks 8 & 9, pp. 207–66.

A close-knit community, like a friendship, has a deeply mutual orientation. In a true community (again, an ideal notion) narrow self-interest is tempered by, indeed transformed by, a concern for the common good—that is, the well-being of the other members of the larger community. As between good friends, individuals within close-knit communities consider acts of self-sacrifice which are beneficial to the common good to be, in part, in their own (self-) interest. Indeed, the similarities between a community, as a group of persons whose own individual good depends on the good of others, and a friendship, leads to the plausible conclusion that a friendship is a kind of community. A friendship is perhaps a small, ardent, and especially intimate community. In this regard John Finnis[6] suggests that "the most intense form of community [is] the friendship of two friends" (*ibid.*, p. 141). Finnis's comparison, which does have some shortcomings, suffices for the present argument. True friendships share enough characteristic features with genuine communities—mutuality, a complex intermingling of self-interest, and some disposition to promote the association itself—that it is reasonable to treat the case for rights between friends as relevant to questions about the compatibility of rights and community.

Now, the problem here is not simply whether persons would enjoy better friendships without having recourse to rights, but whether there is some deep incompatibility between the kind of community promoted by true friendship and the kind of friendship it is possible to coordinate, at least in part, with rights. In short, having—and exercising—rights seems most appropriate in: (1) an individualist atmosphere; (2) an adversarial atmosphere (rights within friendships would be rights *against* friends); (3) an egoistic atmosphere, where sentiments of selfishness run high. Individualism, adversarialism, and egoism are arguably antithetical to the values and the virtues of friendship which are basically communal, cooperative, and often deferential to others. The first claim that the same legal and moral rights held against adversaries hold also between friends seems to lead to a second possibility, that there are in addition even further moral rights specific to friendship itself. But for some of the reasons just outlined this idea of rights specific to friendships also may seem deeply problematic. Finally, what one is able to achieve with rights is clearly not the best that friendship has to offer. For example, to secure help from a friend by a claim of right is hardly to enjoy friendly help. These problems do indicate that while this inquiry may well be relevant to the compatibility of individual rights and the ideals promoted and protected by close-knit communities, the question—Are there rights between friends?—is also of intrinsic interest.

---

[5] *Friendship, Altruism and Morality* (New York: Routledge, 1980).

[6] *Natural Law and Natural Rights* (New York: Oxford, 1980). Aristotle's account of friendship also, of course, links characteristic features of political community and friendship, *Nicomachean Ethics*, bk. 8, pp., 207–32; *The Politics of Aristotle*, Ernest Barker, trans. (New York: Oxford, 1979), bk. 1, pp. 1–38; bk 3, p 120; bk 4, p 181.

II

One criticism of the idea that there are rights between friends has been made famous by Michael Sandel. He argues that friendships, and other close-knit communities, diminish in moral value when they are viewed as susceptible to considerations of justice and claims of rights. Noting that "justice is the first virtue of [only] certain kinds of societies," Sandel suggests we "consider a more or less ideal family situation, where relations are governed in large part by spontaneous affection" and "[i]ndividual rights . . . are seldom invoked, not because injustice is rampant but because their appeal is preempted by a spirit of generosity . . ." (*op. cit.*, pp. 30, 33). Sandel then describes how this ideal community "comes to be wrought with dissension" and the "affection of previous days" gives way to "the observance of rights" (*op. cit.*, p. 33). Sandel's point that "an increase in justice does not necessarily imply an unqualified moral improvement" seems to be, in part, that having *and* observing rights is itself destructive of close-knit community (*op. cit.*, p. 32).

Sandel's overall argument, however, seems to conflate having rights and claiming rights, at least suggesting that claiming rights follows inevitably from having rights. Yet, clearly, having rights does not invariably entail the particular action, or even the general practice, of claiming them. One might have rights without ever claiming them—for one, when they are never violated. One may also have them but not claim them even when they *are* violated—if, in the judgment of the rights holder, making this particular claim, though warranted, is not the most reasonable course of action given other considerations. Sandel is clearly unable to show that merely *having* rights—within the best of close-knit communities, they will likely operate largely as unneeded reminders—is any real challenge to the overall orientation of a community.

Perhaps, then, the crux of Sandel's problem is with the practice of *claiming* rights. The examples he offers of the claiming of rights, however, hardly form a reasonable test case for the evaluation of the effects of claiming rights (or the practice thereof) on close-knit communities. This is because Sandel's paradigm case of claiming rights between friends is, by his own admission, a case of *inappropriate* claiming of rights. When Sandel says "it can also be shown that in some cases justice is not a virtue but a vice," he somewhat suspiciously avoids cases displaying any of the true virtues of rights (for example, providing a response to unjust intrusions). He offers instead an inapposite comparison—misguided claims of right against mistaken foes (*op. cit.*, p. 34). Sandel notes in this regard that "[i]f, out of a *misplaced* sense of justice, a close friend of longstanding repeatedly insists on calculating and paying his precise share of every common expenditure," I shall "feel compelled to be reciprocally scrupulous," and "may begin to wonder whether I have not misunderstood our relationship" (*op. cit.*, p. 35; my emphasis).

Surely, a reasonable understanding of such a case is that one may well *have* misread the nature, or at least the depth, of the friendship in question. The "friend of longstanding" may for various reasons be less friendly than one thought. But the natural conclusion here is not, as Sandel seems to suggest, that the claiming of rights *by itself* was the reason for the distance in the friendship. Rather, it is the friend's decision to use rights in this way (perhaps intentionally "misplaced" to convey dissatisfaction) which has done so. The key point is that in Sandel's example the fact that the claiming of rights was *misplaced* is clearly more important to the dissolution of the friendship than the fact that rights were *claimed*. Likewise, in Sandel's other example when "the harmonious family comes to be wrought with dissension," there is no indication that the claiming of rights is in any way related to the original diminution of family feelings. Here again, it can hardly be said Sandel has shown that claiming rights is a "vice," if there is good reason to claim them. Curiously, Sandel never probes the cases necessary to demonstrate the point he sets forth to make.[7]

---

[7] It is not my point here to deny that claiming rights can ever have a causal effect on the deterioration of a close-knit relationship. On that limited point Sandel is no doubt correct. He has simply not shown—though he seems to assert—that the use of rights in such cases is typically the main cause of unfriendly relations.

Noting the difference between having rights and claiming rights helps alleviate other possible confusions about the existence of rights between friends. One might be said to *have* human rights, or legal rights, against friends without ever choosing, or having good reason, to claim them. Take what are sometimes called *in rem* rights —rights not held against particular individuals, but held, as it were, against the world at large (or the relevant subset thereof). Take, for example, the human right not to be tortured or the legal right to free speech. By noting that one does *have* such rights against all others, this does not exclude one's friends from having the corresponding protective duties. Having justification to *claim* such a fundamental right against a friend would, however, reflect negatively on the friendship. There is every reason to think that Jill's friendship with Jane is somewhat strained if Jill knowingly facilitates those who would torture Jane. Improbable hypothetical scenarios might even be generated here—in a misguided attempt to help, Jill might facilitate Jane's torture at what she supposes would be Jane's request because Jill thinks this may be the best way to save Jane's children. Here even a well-intentioned, though confused, friend can at times be properly constrained by the consideration that *all* persons—even friends—have human rights.

Similarly, Tad's right to free speech binds his friends and enemies alike. All relevantly situated persons, be they friend or foe, have the duty to respect his right. Clearly, Tad's friends, qua friends, are most unlikely to run afoul of this right. This fact does not, of course, entail that his friends have no duties, only perhaps that these are duties which they have no real inclination to avoid.[8] Consider, as well, what are called *in personam* rights—claims where the protective duty is held by a particular person, like the rights deriving from a promise or a contract. Here also, friends may *have* rights against other friends. Indeed, there may sometimes be good reasons for friends to have rights based in promises or contracts, perhaps reasons based principally on considerations of friendship itself. Contracts between friends may at times actually be the most amicable way to protect the closeness of that relationship from the pressures of other less friendly concerns.

III

Now, even if actually having and claiming rights is compatible with the values and virtues of friendship, this does not yet fully solve the problem. One further complication that may be raised by some critics of rights talk is that the very idea that one has rights (not the fact of having them or claiming them) will establish a distance that is destructive of friendship. It may be argued that, like the sense of justice, the sense that individuals have rights somehow puts them on less than friendly terms with other members of close-knit communities. One central issue here is the individualism inherent in the notion of individual rights. The point is that in some fundamental sense the idea that persons have, and may claim, *individual* rights puts the individual at center stage, and thereby inherently at odds with others. In short, this emphasis on the individual's claims has a corrosive effect on the community of friendship. The *idea* of individual rights, however, is compatible with the central values and virtues of friendship.

---

[8] Surely it is wrong to allege (as some readings of Kant seem to suggest) that one does not *have* an obligation simply because one does not in fact *experience* that obligation as a real limitation on, or impediment to, one's present inclinations. There is certainly no *inconsistency* in the idea that one has duties (e.g., some familial duties) that one is really quite happy to carry out.

Various aspects of individualism which are prominent because of this attention to the idea of individual rights deserve mention.[9] First, individual moral rights provide individuals with a certain secure moral standing. As Richard Wasserstrom notes, claiming something as a right "is crucially different from seeking it through the grant of a privilege, the receipt of a favor or the presence of a permission" (*op. cit.*, p. 630). Self-consciously possessed of rights, individuals (even individuals qua friends) can demand that they be treated with a certain (minimal) respect. This, however, does not imply that persons never need favors or indulgence from their friends. It does suggest that in some cases, perhaps where the friendship is suffering trying times, people need not beg or plead with friends for fair treatment. The idea that a person has individual rights—even against friends—provides for a certain powerful sense of self-worth and one typically not provided without individual rights.

This sense of self-worth does not, however, necessarily lead to the use of those rights in ways destructive of friendship. As noted above, a person may *have* rights and even have them violated and yet choose for the moment (for various reasons, including considerations of friendship) not to claim those rights. In this regard, notice also the difference between: (1) actually claiming a right, (2) threatening to claim a right, and (3) noting that you feel a right has been violated.[10] Between friends, the mere observation that one is in a position to claim a right, without yet actually having, or planning, to do so can lead to pointed discussion but not necessarily an adversarial confrontation resulting in the dissolution or diminution of the friendship. One possible characteristic virtue of friendship (and close-knit community) is the mutual willingness to overlook some minor conflicts, or at least the standing inclination to foresee their amicable solution. This virtue, however, does not preclude the very idea that friends have individual rights. It simply suggests one proper framework—a friendly way—for using individual rights, that is, one preserving or even enhancing the friendship.[11] In addition, a proper sense of self-worth (not full-blown, narrowly construed egoism—which will be discussed below) is arguably a precondition for, or at least a *desideratum* of, friendship itself.[12] A proper sense of self-worth is not inconsistent with the virtues of friendship. Clearly, habits of self-respect inspired by rights can encourage such virtues among one's friends.[13]

---

[9] A most useful account of various versions of individualism inherent in rights talk is given by Jeremy Waldron in *Nonsense upon Stilts: Bentham, Burke and Marx on the Rights of Man* (New York: Methuen, 1987), esp. pp 185–8.

[10] Of course, one may relinquish one's rights in various ways (e.g., waiving them). For a similar distinction between asserting and withholding a right and the subsequent effects of doing so on community, see John Tomasini, "Individual Rights and Community Virtues," *Ethics*, CI (1991): 525–7.

A further aspect of individualism that might be seen as destructive of the grounds of friendship is what could be called atomism. In this case, atomism is the view that the individual is isolated from the larger close-knit community by his possession of individual rights. "Liberals" like John Rawls[14] have criticized classical utilitarianism because it "does not take seriously the distinction between persons" (*ibid.*, p. 27). In a sense the problem raised by "communitarian" opponents of rights is that "liberal" individuals take altogether too seriously the distinction—and thereby the difference—between persons. Embracing the ideas of rights, and justice, has led individuals to ignore or downplay their communal similarities. Friends, and close-knit communities in general, value and promote the similarities, not the differences, between individuals. John Hardwig, a feminist advocate of close-knit community, suggests:

> Thinking in terms of rights thus does more than reflect an egoistic, atomistic situation; it *creates* such a situation . . . Because thinking in terms of rights rests on an atomistic picture of us as separate, thinking

---

[11] On the idea that the *way* individuals use their rights is itself a moral issue, see Daniel Callahan, *Setting Limits* (New York: Touchstone, 1988), pp. 176f, my "Dignity, Rights and Self-Control," *Ethics*, XCIX (1989). 526–8; Waldron, "A Right to do Wrong," *Ethics*, XCII, 1 (1981): 21–39.

[12] Cicero holds that a type of self-sufficiency is necessary for friendship. See "On Friendship," in *Other Selves: Philosophers on Friendship*, Michael Pakaluk, ed. (Indianapolis: Hackett, 1991), pp. 79–116.

[13] This is not, of course, to imply that friendship is not itself a powerful source of a sense of self-worth. It is The point is simply that a person's sense of self-worth is quite complex and has various sources.

[14] *A Theory of Justice* (Cambridge· Harvard, 1971).

> in terms of rights systematically denies the unity, the togetherness, the "we" that we are trying to create (*op. cit.*, p. 449; my emphasis).[15]

In light of such an argument, it is worth recalling that even the closest of friends are in fact separate in some important ways. Indeed, some recognition of these differences is valuable to understanding fully the ideal of friendship itself. Friends, of course, inhabit separate bodies. Each friend also has unique perceptions and ideas, attitudes and emotions. These simple facts alone account for some significant differences between even very close friends. On occasion good friends differ over fairly basic matters: what each believes is beautiful, amusing, exciting, perhaps even sometimes what is required by morality. Most poignantly, friends may even disagree about what is called for by friendship itself.

It is also true that friends are in many ways quite similar in their experience and evaluation of the world. Aristotle suggests that in the highest form of friendship the virtues of the good man are mirrored back to and reinforced by his friends.[16] Yet what makes this experience of mutual support especially valuable for friends is the recognition and enjoyment of the fact that each individual person is part of that friendship as a distinct, unique person. Leaders with sycophantic advisors tend to get less than the best advice. Having one's views reinforced by someone who cannot do otherwise—very young children are at times related to their parents this way—is, in truth, considerably less reassuring than genuine autonomous support.

As friends, Jane and Jill may well be more similar, in some fundamental ways, than they are different. Yet their distinct individuality adds to, rather than detracts from, the quality of their friendship. It may even enhance their experience of that friendship. In a world of strangers, friends are especially esteemed because their differences from us do not estrange us from them. These inevitable differences between friends also shape the process of becoming friends. The gradual integration into any friendship proceeds from the fact of prior differentiation. Deepening friendships progress with the growing appreciation that the friend has considerable life experience prior to, and outside of, the friendship. In general, discovering friendly similarities in other individuals does not require ignoring the very real differences within close-knit relationships.[17]

---

[15] For one account of why some feminists have a similar antagonism to rights as "an idiom associated with masculine concerns," see Waldron, *Nonsense upon Stilts*, pp. 159–60.

[16] *Nicomachean Ethics*, bk. 9, esp. chs. 4 & 8, pp 245–8, 253–6

IV

Finally, the idea of individual rights will be seen by some as an independent source of *egoism*. From the point of view of friendship, the divisive passion characteristic of egoism is selfishness. For present purposes selfishness, or *narrow* self-interest, can be characterized as an overriding concern for one's own self-interest without any (or proper) concern for how this affects others. Jeremy Waldron notes that Bentham claims "that the encouragement of the selfish passions is 'the perpetual and palpable *object* of a Declaration of Right'."[18] But is crass selfishness a product of the very idea that individuals possess rights? While the *use* of rights in any particular case may not be wise or prudent, Waldron notes that:

> A theory of rights leaves agents needing more in the way of moral guidance, *not* because rights are egoistic . . . but because rights offer no guidance at all—egoistic or otherwise—to those who have them . . . Certainly rights aim to weaken some of the *bonds* of social life, but that is to be construed as the promotion of egoism only by those who consider community essentially *bondage* . . (*ibid.*, p. 195)

Waldron's observation helps explain how the idea of rights between friends may avoid the various pitfalls of an unfriendly selfishness.

The idea of individual rights allows for and in some cases may require a discussion of their proper, or even their optimal, use in particular cases. The use of rights can be, and between friends should be, guided by what might be called communal values. For example, one of the values protected by rights between friends would be mutual trust. Tad has a right that Tim not furtively read Tad's private journal even though Tim may quite innocently want to know his good friend even better. This right protects the enjoyment of a range of values. It protects Tad's individual privacy. It also helps secure the ongoing trust between them as friends.

---

[17] A. I. Melden, *Rights in Moral Lives* (Berkeley: California UP, 1988), pp. 124, 126, is one of a few recent writers at least to mention rights between friends; cf. also Neera K. Badhwar, "The Circumstances of Justice: Pluralism, Community, and Friendship" (forthcoming in *The Journal of Political Philosophy*, I). Henry Sidgwick, *The Methods of Ethics* (London: Macmillan, 1907), pp. 248–50, 256–9, has an insightful discussion of *duties* (if not yet rights) between friends, as does Kent Greenawalt in "The Right to Silence and Human Dignity," in *The Constitution of Rights: Human Dignity and American Values*, Michael J Meyer & William A. Parent, eds. (Ithaca: Cornell, 1992), pp. 192–209

[18] Waldron, *Nonsense upon Stilts*, p. 191; the entire quotation from Bentham is on p. 48. Waldron has a good account of related senses of selfishness, pp. 191–201.

We can distinguish, roughly, between personal values (states of affairs valued independently of one's relation to others—e.g., individual physical security) and communal values (those valued because of one's membership in some larger community—e.g., mutual trust).[19] An egoist, having little or no concern for how his self-interest affects others, will be unable to appreciate communal values, except perhaps instrumentally. As a good friend, however, Tim would value Tad's privacy for Tad's sake. While an unfair invasion of Tad's privacy by a third party would hurt Tad, it would thereby, as his friend, hurt Tim as well. But Tim's interest in proper respect for Tad's privacy is also rooted in his concern for mutual trust within the friendship. The trust destroyed by *Tim's* disrespectful invasion of Tad's privacy will likely be a much deeper violation than that of any stranger. This violation hurts both Tad and Tim; it also hurts their friendship. It damages an essential mutual trust. To suggest that this disrespectful invasion of privacy harms their friendship is to say that it harms Tad and Tim *as friends*. As a friend one has an individual stake in another's personal values. One also has a stake in the communal relationship itself—that is, in part, one has a stake in one's own position as a friend.

Yet Waldron's claim that there are "many *types* of human relation where standing on one's rights *would* be inappropriate" seems overstated.[20] All that is proper to note here is that there clearly are *times* within some close-knit communities where it would be wrong—that is, inappropriately egoistic or in some way misguided—to do so. After Tad and Tim have agreed to share private journals, and after Tad has already shared his own, then it might well be wrong for Tim to raise the issue of his rights to privacy.[21] The simple point is that friendships should not be viewed as a special kind of relationship which eschews rights talk. Rather, friends require a special sensitivity to the *reasons* provided for various uses of rights.

---

[19] For interesting but more elaborate accounts of "communal values" than it is necessary for present purposes to explicate, see Gerald J. Postema, "Collective Evils, Harms and the Law," *Ethics*, XCVII, 2 (1987): 414–40; and Mark Sagoff, "Values and Preferences," *Ethics*, XCVI, 2 (1986): 301–16.

[20] Waldron, *Nonsense upon Stilts*, pp 188–9 (my emphasis on 'types'). It is unclear why Waldron makes this point given that he says (pp. 191–2) that to have a right is neither to be protected from all moral criticism nor a seal of moral approval.

[21] Even here it may be a difficult case The problem is not that Tim has no rights to privacy because he waived all of them as part of his agreement to share journals If Tim comes to a reasonable decision that by sharing his journal Tad has not revealed himself at all—perhaps his journal is especially prosaic and reticent—Tim may be tempted to reassert that his own right to privacy exempts him from the claim against him that he share it with Tad. The actual assertion of such a claim may well compromise the friendship with Tad, but if it comes to that, Tim may already have been put off by Tad's meager (or manipulative) attempt at openness

Many of these aspects of individualism outlined in the last three sections, and thought to be inherent in the idea of individual rights, may appear to culminate in the prospect of *adversarial* relations between friends. Rights between friends are, of course, rights held *against* friends. Yet the potential adversarialism inhering in the idea of rights exists only insofar as their actual use has that intent or outcome. As argued above against Sandel, the adversarial use of rights is not an inevitable consequence of having rights, neither is an adversarial posture somehow inherent in the idea of rights. Now clearly, a theory of what rights one has should not be taken as a complete account of one's moral relations. Equally clearly—but less often recognized by proponents of individual rights—an account of individual rights should be supplemented by a theory of virtue. This is necessary because the idea of oneself as possessed of certain rights falls short of thoughtful guidance on their use in specific circumstances. The idea of rights would be greatly benefited by an account of their reasonable use in particular contexts. Claiming rights can be modulated by a commitment to many values, not the least of which are those of friendship and community.[22] In this way it can be appreciated that the *idea* of individual rights between friends need not be the path to unfriendly individualism.

V

One important final question exists about the character, or the quality, of any friendship that one regulates, in part, with rights. It might be said that the essence of some communal values, like mutual trust between friends, cannot be maintained by claims of right. The problem can be put this way: what friends want out of any friendship is, in its very nature, diminished or destroyed, if it is secured through a claim of right. In short, *from a friend* one wants spontaneous, not grudging, concern. The quality of the concern received, if one can claim it as a right, is simply not the same as if it is spontaneously given.[23] Or, more strongly, if concern is got by claiming it, the essence of what is really desired—spontaneous care—is simply effaced. Two points about this observation are worthy of attention.

---

[22] Waldron, *Nonsense upon Stilts*, p. 196.

[23] This point was first pointed out to me in this context by Badhwar. On this question of spontaneity, cf. her "Circumstances of Justice: Pluralism, Community, and Friendship," pp. 27–9 (in manuscript) See also Bernard Williams, "Persons, Character, and Morality," in *Moral Luck* (New York: Cambridge, 1981); and Marcia Baron, "The Alleged Moral Repugnance of Acting from Duty," this JOURNAL, LXXXI, 4 (April 1984): 197–220.

First, to say that a friendship is regulated, in part, with rights is not to say that every aspect of the friendship is a matter of claiming rights. It has not been my argument here that rights are always the only, or the best, way to regulate a close-knit community. A very good friendship will transcend the need to *claim* rights most of the time. And a perfect friendship might well always forego the claiming of rights. But not all friendships—sometimes perhaps not the best of them—are perfect. Even between close friends of longstanding, avoiding certain potential transgressions may require respect for the friend's independent status, and at least her possession of some rights. If one fully recognizes the independent status of a friend, then she will not be seen primarily as someone living to fulfill all of one's own desires. In the best of friendships there is no direct harm to the friendship, or the quality of the concern thereby derived, to see the other as a bearer of rights, since this status alone entails no actual claims. Here, spontaneous concern is not destroyed by *claiming* it as a right, because no actual claim is ever made.

Second, to note that receiving respect by a claim of right (or out of regard for the possession of rights) is not the best that friendship has to offer is not yet to show that such respect is no good at all, even between friends. Jane may feel that, if in some cases Jill respects her privacy only after it is claimed as a right, their friendship has a long way to go to suit Jane. Of course, that is not to say that Jane would rather have her privacy violated than claim respect for it. Neither does it show that in such a case Jane would necessarily wish to have never become, or no longer remain, Jill's friend. Even between friends it is sometimes better to enjoy certain goods, like respect, by claiming them, than to suffer their loss *or* to suffer the loss of the friendship.[24] Securing respect between friends, even as an afterthought of sorts, can be better than not securing it at all. The obvious desirability of faultlessly spontaneous respect and concern notwithstanding, from the point of view of friendship, second best is not utterly unacceptable in every case. What clearly *is* unacceptable between friends is a steady diet of antagonism and selfishness. While it seems that only unusual friendships—if any at all—survive persistent egoism, rights between friends, properly understood, militate against egoism. Even the most well-meaning of friends may, at times, fail to accord proper respect to a friend. Any view of friendship that rules out the very possibility of insensitive disrespect or conflict between friends seems at best a romantic, at worst a wholly unrealistic, view of friendship.

---

[24] What may happen in such a case is that between friends the scope of a right to privacy sometimes becomes less clear. This may also be a case where a right need not actually be claimed to secure the respect in question. While even this compliance because of the recognition of but not the claiming of rights falls short of the spontaneous respect often expected from friends, it goes too far to say that here the respect is *destroyed*. It is diminished, but diminished respect (even from friends) may be better than the loss of the respect or the friendship.

VI

If there are rights between friends—especially rights specific to friendship itself—What are they? I shall mention only a few rights here so as to initiate this discussion. These could be taken as possible paradigm cases of rights that hold between very close friends. In addition, what would be taken as a good reason to claim such rights in individual cases is, in part, a function of the character, the depth, and the nature of the particular friendship. This being the case, little can be said outside of more specific contexts about when and how it would be reasonable, or reasonably friendly, to exercise such rights. Finally, nothing argued so far should be taken to show that one has a right *to* a particular friendship (or, for that matter, friendship in general). The rights delineated here arise out of existing friendships.[25]

I have already discussed rights to respect and concern between friends, especially respect for some level of privacy between friends. But there is nothing unique to friendship about privacy (that is, one expects to be accorded some privacy—likely even more—from strangers). So, take as a further example a right that is arguably unique to friendship and the close relationship it fosters. Between very close friends there is a right to be treated openly within the relationship. It is important to distinguish this claim to be treated openly from the claim to be treated honestly. Although openness and honesty sometimes have mutually reinforcing effects, openness goes beyond a simple willingness to tell the truth. It is a disposition to be forthright about it. Among close friends the virtue of openness is a general disposition to be forthcoming with some of the truth, especially about oneself. One is hardly expected, in this sense, to be open with strangers or even mere acquaintances. But some willingness to communicate openly one's own concerns, hopes, joys, and fears is essential to the type of sharing one comes not only to desire but to expect—especially when one is open oneself—from very close friends. Such openness is not, however, to be mistaken for a confessional disposition. To meet legitimate expectations of mutual openness, Jane need not always bare her soul to Jill. The desirability of this one aspect of intimacy does not entail that even very close friends need be intimates at all times or in every way. But certainly the ongoing deepening of a close friendship requires, from the point of view of each individual friend, a commitment to some openness and intimacy.[26]

---

[25] To say that such rights derive from the commitment (at least tacit or hypothetical) already made within existing friendships is obviously not a full account of how such special rights arise. That topic must be the subject of another paper

Between close friends one also has a right to concern and respect for *one's own sake*. A general disposition to respect such a claim seems to be another virtue of friendship—the virtue of mutual altruism. This regard for one's friend for her own sake, and not for the sake of someone else, is also one of the core values of friendship. To appreciate this particular claim between friends one need only recognize the strong sense in which one should regard one's true friends as unique. In other words, it violates the proper respect due to a close friend simply to replace her friendship with the friendship of someone else (to believe that one's friends are fungible, easily exchanged like a tourist swapping foreign currencies). Similarly, it violates any friend's basic rights to value her in order to enhance a friendship with someone else. None of this is to say that it is inappropriate to have a wide range of friends or share different aspects of oneself with different friends. It is simply to note that it is too weak to say that Jane would be put off if she discovered Jill was her friend primarily as a way to get to know Tom better. Jane, and her claim to basic respect as a friend, would be violated in a deep way if Jill was her friend not primarily because of anything about Jane, but for the sake of Tom's friendship. True friends have a right to be treated not as simply useful to other goals, or as generic but fungible placeholders, but as real and unique people. Part of what is deeply important about friendship is being cared about because of (and sometimes in spite of) the singular and distinctive person that one is.

Between close friends other rights include a right not always to be asked to place a friendship before one's other needs, including other friends. Friends have a right that their closest and most friendly relations not be all consuming *all* of the time. Between close friends one also has the right to the benefit of the doubt in cases of conflict. This is not to suppose that friends are never inappropriately egoistic but only that a close friend has a right that this not immediately be thought to be the cause of a conflict. The observance of the existence of such a right helps put conflicts between friends in an appropriate perspective. Between close friends each person also has a right to roughly equal influence on the friendship. Observing such a right reflects an essential commitment to compromise. In fact, the spirit of this virtue of compromise underlies the likelihood of the general observance of all the rights between friends mentioned so far. Whatever kindness one shows to a close friend, one can usually expect him to be disposed to return it in full. Of course, such equal influence is never perfect, but failing to honor the claim to it not only harms the friendship but fails to accord proper respect to the individual friend in question.

---

[26] It is not my point here that every failure of openness or intimacy is best described as a violation of rights, but only that a complete and unilateral failure to be open might well be so described.

In conclusion, it must be emphasized that any account of rights between friends is clearly not all that there is to friendship. Not everything one has reason to give to friends is owed as a matter of rights. This indeed would be an impoverished view of friendship. It has not been argued here that friendship is fundamentally right-based—that is, neither the practice of friendship nor the basic moral impulse that gives rise to it can be reduced to the observance of rights.[27] Close friends will spontaneously go to great lengths for one another; and they will do so without either claiming rights or giving any thought to the existence of rights. But this happy fact about friends ought not tempt us to embrace a utopian account of friendship which puts friends beyond the *possibility* of conflict. Clearly, such a fanciful vision would fail as a suitable account of actual friendships and close-knit communities.

Equally clearly, it is deeply misguided to suggest, as some opponents of rights in close-knit communities do, that "if you are my friend . . . [y]ou can invade my privacy, . . . fail to respect my private property, verbally abuse or perhaps even physically assault me, and it is all right so long as I know that you are my friend."[28] No individual need mutely suffer harm and abuse from strangers or friends. The virtues of self-respect offered by individual rights need not take a back seat to the excesses of some close-knit relationships. The fundamental mistake of seeing the rights of friends as antagonistic to the virtues of friendship obscures our vision of the compatibility of individual rights and close-knit community. Even beyond this compatibility, thinking about rights between friends may help us both understand something about the place of individuals within close-knit communities, and clarify the idea that there are reasonable, and unreasonable, uses of some rights. At times to protect themselves, and the friendship, individual friends—qua friends—must be positioned to go beyond merely explaining what they need. Rights provide one good, and sometimes proper, way of doing so. If, between friends, we have reason to take rights seriously, that does not mean we take friendship any less so.

MICHAEL J MEYER

Santa Clara University

---

[27] Nor has it been claimed here that the virtues of friendship are of instrumental value only—that is instrumental to the protection of rights See Joseph Raz, "Right-Based Moralities," in *Theories of Rights*, Waldron, ed. (New York Oxford, 1984), pp. 182–200.

[28] Hardwig, p 56

RATIONALITY AND THE PRISONER'S DILEMMA IN DAVID GAUTHIER'S *MORALS BY AGREEMENT*[*]

In *Morals by Agreement,* David Gauthier[1] defends "the traditional conception of morality as a *rational constraint* on the pursuit of individual interest" (MA, 2; my italics). His defense is set against the background of a prominent view of rational conduct, the view that Gauthier calls "the utility maximizing conception of rationality" (MA, 6f). The maximizing conception in its standard interpretation holds that

> an act of a person *p* is rational if and only if it is expected to maximize *p's own* interest or preferences.[2]

This principle codifies the idea of immediate connection between rationality and the direct pursuit of individual interest—the idea that "there is a connection between [practical] reason and interest —or advantage, benefit, preference, satisfaction, or individual utility" (MA, 6).

The traditional conception of morality Gauthier attempts to defend is clearly at odds with the maximizing view of rationality. The conception of morality does not deny the apparent fact that morality may and often times does conflict with the direct pursuit of individual interest; and it holds that in some cases of such conflicts it is rational to comply with moral commands at the cost of one's own interest or preferences. Thus, its defense requires justifying why it is rational at all to comply with moral commands even when they require sacrificing one's own interest.

Yet Gauthier wants to base his argument on the maximizing conception of rationality itself. He somehow thinks that the maximizing conception, in one version or another, is "the only one capable of critical examination" (MA, vi). Whether or not one agrees with this, his attempt, if successful, would result in a strong and desirable defense of morality. The opponents of (the rationality of) morality typically adhere to the maximizing conception;[3] so the justification

---

[*] I am grateful to David Gauthier, Gregory Kavka, and Philippa Foot for their comments on the earlier versions of this paper.

[1] New York: Oxford, 1986. Hereafter MA.

[2] Cf. MA, 6 "a person acts rationally if and only if she seeks her greatest interest or benefit."

[3] In particular, I take Gauthier's opponent, Hobbes's Foole (cf. fn. 8), as an avowed promulgator of what he takes to be the correct implication of rationality as maximizing of individual interest.

0022-362X/92/8909/484–95

of morality in terms of their own conception of rationality, if available, will turn out to be the most conclusive argument against them.

The strategy of his argument might be best explained with an analysis of the prisoner's dilemma. Recall a standard description of the (two-person) prisoner's dilemma, the possible outcomes of which are summarized in the following table, where numbers indicate the years the two suspects, he and I, will live in jail:[4]

		He	
	(I, he)	Confess	Not
I	Confess	(5, 5)	(1, 10)
	Not	(10, 1)	(2, 2)

Table 1

In this situation, adherence to the maximizing rationality by both agents—i.e., their unrestrained pursuits of their respective individual interests—yields them a suboptimal outcome: 5 years in jail for each. If, however, they can effectively cooperate, they can reap a better outcome of two years in jail for each, an outcome which is fair and more advantageous to each than that of their individually "rational" actions.

Gauthier's approach is to view morality as a code agreed upon to enable a fair and mutually advantageous social cooperation, which, as the prisoner's dilemma is apt to show, may sometimes be at variance with the direct pursuit of individual interest. To justify the constraint that the contractarian morality imposes on the direct pursuit of individual interest, Gauthier argues that when the agents in a situation characterized by "a conflict as well as identity of interests" (*ibid.*, p. 126) (such as the prisoner's dilemma) attempt to reach the optimal cooperational outcome on the basis of the maximizing conception of rationality, they find it rational (in the maximizing sense) to agree not to confess. *Furthermore,* he argues, they find it rational to form a disposition to abide by the agreement so made—given certain provisos. That is, straightforward pursuers of individual interests ("straightforward maximizers") have the reason (by their own standards) to change into "constrained maximizers," namely, those who restrain the pursuits of their respective interests in terms of, say, the contractarian morality.

---

[4] Here I shall not describe the prisoner's dilemma I just refer to the description in MA, 79–82. See also John Rawls, *A Theory of Justice* (Cambridge: Harvard, 1971), p. 269.

In order to elucidate the nature of his argument, I shall follow Gauthier in concentrating on the analysis of what it is to act rationally in the prisoner's dilemma. There I shall attempt to show why the rationality of (the act of) forming a disposition is one thing and the rationality of acting in conformity with the disposition quite another. Thus, I argue, in justifying the rationality of complying with morality viewed as a code enabling social cooperation, Gauthier has as yet to vindicate the principle connecting the two:

> If it is rational to be disposed to restraint in the direct pursuit of individual interest, then it is rational to act in accordance with the disposition.

Finally, reflecting on the implications of adopting this principle, I shall argue that Gauthier's justification, properly understood, turns out to *presuppose*, rather than to derive, that there exists an implicit constraint on the direct pursuit of individual interest in what he should argue is the proper interpretation of "the maximizing conception" of rationality.

I

Stage I. Recall table 1 of payoffs proposed to two prisoners, he and I. Given the maximizing conception (in its standard interpretation), it is rational for me to confess under that circumstance. In this context it is convenient to talk in terms of the expected years during which I shall live in jail depending on whether I choose to confess or not—assuming that the longer the expected years, the less the expected utility (alternatively, the less it is in my interest or the less I prefer it). Now, depending on whether or not I confess, the measure, Y, of expected years in jail may be calculated as follows:

$$Y\text{(confess)} \quad = 5 \times p + 1 \times (1 - p)$$

$$Y\text{(not confess)} = 10 \times p + 2 \times (1 - p)$$

where p $(0 \leq p \leq)$ is the probability that I initially assign to the chance that he will confess. Now, Y (not confess) is greater than Y (confess) whatever the initial probability, p, of his confessing may be;[5] for $10 \times p$ is greater than $5 \times p$ and $2 \times (1 - p)$ is greater than $1 - p$. That is, confessing is rational for me.

I also realize that the situation is symmetrical between him and me; given that he is a rational agent with a perfect power of reasoning, it is also clear to me that he, too, will choose rationally to confess. Consequently, my reasoning continues, I am doomed to live 5 years in jail; that is, in this situation the measure of the expected years in prison is 5 *when everything,* including his rational action, *is taken into account.* Thus, I naturally want to change the situation so as to make him not to confess; and it is the same with him. But, alas, that option is not available to us; for we cannot coordinate our actions by way of, say, communication.

---

[5] One may reason that whether or not he is assumed to confess I shall be better off if I confess This reasoning takes into account the two limiting cases when p is respectively 0 and 1

Stage II. Suppose now that the situation has slightly changed; that is, we are allowed to communicate once. In the meeting I initially try to dissuade him from confessing, but I shortly find out that I can dissuade him only if I also promise not to confess and vice versa. Consequently, we make an agreement not to confess. Assuming, following Gauthier (MA, 156), that the bargaining is cost-free, it seems rational to make such an agreement; for it renders my expected utility *at least as high as* otherwise.[6] In fact, the actual discussion might not have been essential. Suppose that, fortunately, we previously read Gauthier's book together and that, therefore, we were convinced of, say, Gauthier's principle of "minimax relative concession," which specifies terms of a fair and mutually advantageous cooperation.[7] Furthermore, each of us knows that the other has the same understanding. Then I can conclude that we *would have agreed* not to confess were there a chance to discuss. And I, knowing his acumen, can be certain that he should be reasoning in the same way. In sum, it is exactly as if we made an actual agreement; so we might rely on the hypothetical agreement.

Does the agreement, hypothetical or actual, make a substantial difference to the situation? Here Hobbes's Foole[8] enters into the stage, arguing that it makes no essential difference to the situation. Given the maximizing conception, the Foole argues, the agreement falls short of introducing an intrinsic reason to comply with the agreement. Consequently, he continues, the agreement is not sufficient to secure the cooperational benefits to us, although the agreement seems rational to make and is *necessary* for cooperation; for I still have a choice between complying with and deserting the agreement. Here is the table of possible outcomes:

---

[6] I shall argue that in the end the agreement fails to upgrade my expected utility Thus, as Kavka correctly noted, I as a perfectly rational agent may not find it the only rational action to make the agreement since I can see that it is pointless. (Insofar as maximizing rationality is concerned, on the other hand, it is not irrational, either, to make the agreement given that the bargaining is cost-free.) Yet the major point in this section remains valid: even given the agreement, it is still rational, in the maximizing sense, for me to confess.

[7] For Gauthier's account of the principle of minimax relative concession, see MA, 141–5. In this context, however, the specific content of the principle is not crucial. What is important is that we both recognize the optimal outcome as the one based on a/the fair and mutually advantageous principle.

[8] See Hobbes, *Leviathan,* C. B. Macpherson, ed. (New York: Penguin, 1968), ch. 15, p. 203: "The Foole . . . does not therein deny, that there be Covenants; but he questionth, whether Injustice . . . may not sometimes stand with that Reason, which dictateth to every man his own good "

		He	
	(I, he)	Not	Comply
I	Not	(5, 5)	(1, 10)
	Comply	(10, 1)	(2, 2)

Table 2

It shows that I am facing the same situation as the original dilemma; so I find it rational not to comply with the agreement, i.e., rational to confess. The agreement turns out to be of no avail, although it initially seems rational to make. The Foole, as Gauthier notes, is challenging "what would seem to be the ordinary view that *given neither unforseen circumstances nor misrepresentation of terms,* it is rational to comply with an agreement if it is rational to make it" (MA, 165; my emphasis).

Stage III. My dilemma persisting in stage II lies in the fact that I did not succeed in making him not confess even by reaching an agreement with him; consequently, the agreement fails to upgrade my expectation. What is needed is not making an agreement but ensuring his compliance.

Suppose now that we are granted another chance to meet. Before the meeting I am considering how to make it certain that he will comply. To be sure, the best result of the meeting would be to ensure his unilateral compliance regardless of my action, but this is simply impossible. So I look for a way of achieving the outcome that is the second best result for me: making it certain that he will comply by way of ensuring him of my own compliance. Otherwise, I have no way to shorten 5 years in jail. The way to ascertain my compliance is to form and make him check, with absolute certainty, my disposition to constrained maximization, i.e., the disposition to comply with an agreement based on the fair and mutually advantageous principle in case I am certain that he will also comply with it.[9] He reaches the same conclusion on his own, and both of us succeed in forming the requisite disposition. In the meeting we "see" each other's disposition, and reach again or confirm the agreement not to confess.

---

[9] This is a simplified version of the disposition to constrained maximization, the version which I think is adequate for the simplified situation at issue here. Cf. MA, 167f.

Now I can reliably expect to gain the benefit of 2 years in jail. By forming the disposition, I have succeeded in lowering the ultimate measure, *Y*, of expected years in jail to 2.

II

In stage III, the dilemma facing the prisoners is given a practical solution; they succeed in reaping the benefit of cooperation through an agreement backed up by the disposition to constrained maximization. And it is rational in the maximizing sense to be so disposed in such a situation. Yet, I argue, the problem of justifying the rationality of acting as the disposition leads one to is far from being solved.

In my analysis of the solution of the dilemma in stage III, it is not necessary at all to assume that it is rational for him or "me" to act in accordance with the disposition to constrained maximization; there it is sufficient to observe that the prisoners (he and "I") are now reliably expected to act in accordance with the disposition—that is, not to confess—because the disposition is firmly entrenched. In their ("our") own deliberation, too, their expectations about the other's action are entirely based on the available information concerning the other's dispositions. Although it is argued that it is rational to form the disposition to constrained maximization, no appeal needs to be made to the rationality of acting in accordance with the disposition. In sum, the rationality of internalizing a fair and mutually advantageous principle of cooperation is not yet shown to imply that it is rational to act in accordance with it.

Thus, Hobbes's Foole, the staunchest promulgator of the maximizing conception (in its standard interpretation), will come back to the stage and continue to argue against Gauthier's defense of morality as follows. Consider again, the Foole argues, the situation where we have found each other's disposition to constrained maximization with absolute certainty. I, the Foole, grant that it was rational to be so disposed. But the question remains whether it is rational to act in accordance with the disposition—to be sure, granting that the agreement is fair[10] and mutually advantageous, and that neither unforseen circumstances nor misrepresentation of terms is found, and that "the rationale for [the disposition] does not appeal to any weakness or imperfection in [our] reasoning" (MA, 186).[11]

---

[10] I understand this as including fairness of the initial position of the agreement, which Gauthier argues is ensured by satisfaction of what he calls "the Lockean proviso."

[11] All these assumptions are satisfied by description of the situation itself.

There are indeed two kinds of differences between the present situation (i.e., stage III) and the dilemma in the stage II. On the one hand, there are differences based on the fact that I am myself disposed to constrained maximization; on the other hand, those deriving from the fact that I now have different expectations about the other's action because I know that he has a firm disposition to constrained maximization. The first kind of differences, I argue, do not count when the *rationality* of the conforming action is at issue. The rationale for this consists in the distinction between questions concerning rationality and those concerning motivational strength. I, the Foole, concede that by forming the disposition I made myself strongly disposed not to confess in the present situation. Consequently, I *know* that I *will not* confess. But this is a far cry from saying it is rational to act so.

Imagine that I am reflecting on the situation several years later after not confessing: I came out of the jail after 2 years with complacence in the successful cooperation. At this point, there is no longer any doubt about which motivation *was* the strongest when I was deciding whether or not to confess. But the question still remains whether it was rational (by the standard of that time) to act in accordance with the disposition, and the answer is not automatically affirmative. We often meet with another kind of situation in which the distinction is typically made between questions concerning rationality and those concerning motivational strength. The fact that I failed to act in a certain way is not automatically regarded as implying that it was irrational for me to act in that way; the discrepancy is often explained with an appeal to what is called "weakness of the will." If this is a real phenomenon, why should motivational strength of a certain disposition directly imply the rationality of acting in accordance with it, even granting that the disposition itself was formed on a rational basis?

Then what is rational for me to do in stage III? I, the Foole, again suggest a simple calculation of the expected utilities. Look at table 3 below:

		He	
	(I, he)	Not	Comform
I	Not	(5, 5)	(1, 10)
	Conform to the disposition	(10, 1)	(2, 2)

Table 3

Table 3 is exactly like table 2, but in calculating expected utilities I may also take it into account that I now expect him to act in conformity with the disposition. Consequently, only the second column in table 3 really matters. Here are the measures, Y, of the expected years in jail, depending on whether I also act in conformity with the disposition or not:

$$Y\,(\text{Not conforming}) = 1$$

$$Y\,(\text{Conforming}) = 2$$

This shows that not conforming, i.e., confessing, is rational in spite of my entrenched disposition not to confess.

To emphasize again, this does not mean that I shall not act in conformity with the disposition. Rather, I take this as a situation in which I absolutely feel like not confessing, although I am aware that it is not rational. That is exactly what he, the other party, wanted to make certain; it did not concern him at all whether I would find it rational to conform to the disposition once it was made absolutely certain that I would conform.

III

I presented the Foole's argument as consisting of two parts. In the first part, he argues that the following principle is yet to be justified:

1. *Connecting principle:* (given that typical provisos[12] are met) if it is rational to be disposed to constrained maximization, then it is rational to act in conformity with the disposition.

The latter part consists in persistently applying the maximizing conception of rationality to the action of conforming to the disposition, which leads to the conclusion that confessing is still rational. To be sure, Gauthier agrees with neither of these. I shall begin by discussing his disagreement with the first point. The discussion will eventually lead to his rejection of the second point.

---

[12] For some of the provisos, see above.

[13] I understand Gauthier's perfect actor as the actor who is without "any weakness or imperfection in the reasoning." Note that the Foole's argument for the first point takes no advantage of such imperfections. On the other hand, if he means by a perfect actor a person who chooses to do something if and only if it is rational to do so, then he begs the question by assuming that the perfect actor will choose to act in conformity with the rationally grounded disposition.

Gauthier might respond that in his analysis of the prisoner's dilemma the agents are assumed to be perfect actors in the latter, stronger sense. It is indeed true that the "straightforward maximizers" are perfect actors, given the maximizing conception. Yet the issue is whether they remain to be so when they change themselves into "constrained maximizers." This is not ensured by the rationality of the metamorphosing actions The Foole's straightforward answer, to be sure, is in the negative.

He affirms that "if her [the perfect actor's] dispositions to choose are rational, then surely her choices are also rational" (MA, 186).[13] By this he seems to endorse principle (1). But the view itself needs to be justified, and the justification should be based on (some interpretation of) the maximizing conception of rationality; otherwise, he has yet to make his case that he is giving "a utility-maximizing rationale" (MA, 158) for compliance. Principle (1) might be granted to be part of our common conception of rationality, but that is another matter.

Imagine that principle (1) might be justified by the maximizing conception of rationality as it is formulated in the beginning of this paper:[14]

2. *Action-maximizing principle:* an act is rational if and only if it is expected to maximize the agent's own interest.

Or, suppose that it is justified as a kind of conceptual truth linking the concept of the rationality of a disposition to that of the rationality of conforming action or that it is justified, independently of principle (2), as reflecting a fixed point in our intuitions about the concept of rationality. Then it is rather worse than better for principle (2) because it results in a *reductio* of (2): by utilizing Gauthier's argument for the rationality of being disposed to constrained maximization and the implication of the principle (2) drawn by the Foole, one can show that, given relevant assumptions built into the description of the prisoner's dilemma, principle (2) together with principle (1) leads to a contradiction. Let me explain.

Here is the conclusion of Gauthier's argument, which I grant: in the prisoner's dilemma,

3. It is rational to be disposed to constrained maximization.

(1) and (3) imply, as is intended, that

4 It is rational to act in conformity with the disposition to constrained maximization.

Now, apply principle (2) to the stage III quasi dilemma. Then, as the Foole is apt to show, it follows that:

5. It is rational not to act in conformity with the disposition to constrained maximization.

and that:

6. It is not rational to act in conformity with the disposition.

---

[14] It was based on Gauthier's formulation in MA, ch. 1. Cf. fn. 2.

which is due to the "only if" part of principle (2). (6) contradicts (4). The prisoner's dilemma, it turns out, helps to elucidate that principles (1) and (2), each of which taken independently seems adequate, lead to a contradiction when applied to the realistic situation marked by "conflict as well as identity of interests."

Gauthier may be anticipating this problem when he says that he is working with a different interpretation of the maximizing conception, which "identifies practical rationality with utility-maximization at the level of dispositions to choose." And he "carries through the implication of that identification in assessing the rationality of particular choices" (MA, 187). Thus, while rejecting (2) he might be countenancing the following principle:

7. *Principle of disposition maximizing:* a disposition is rational for a person *p* to have if and only if it is expected to maximize *p*'s own interest.

In addition, he certainly needs principle (1), whatever its justification may be, to show the rationality of acting in conformity with the disposition to constrained maximization.

It is quite obvious that principle (1) should be independent of the disposition-maximizing principle since this has nothing whatsoever to say about the rationality of particular actions. Gauthier might continue that it is so much the better for his disposition-maximizing conception of rationality. Principle (1) may be simply added as an independent principle to the conception. One might motivate the move in the following way: we begin by reflecting on the maximizing conception in its inarticulate form, and arrive at a correct formulation of it; and then we can appropriately extend that conception to reach a coherent and comprehensive concept of rationality. Now think about the present situation: the Foole's persistent argument makes it clear that principle (1) needs to be incorporated into the concept of rationality to launch Gauthier's justification of morality, and it turns out that adding it to the disposition-maximizing principle can have no bad consequences; then, it is rather mandatory to add it (or some other principle that implies it). Or so he might argue. This might exemplify a respectable methodology, but it is still worth noting that he is, in effect, implicitly working with at least two independent principles of practical rationality to show the rationality of morality; he no longer undertakes to derive the one from the other.

Finally, insofar as he wants to endorse the "maximizing conception" in its correct interpretation, he will certainly agree to adding the following principle:

8. An action is rational if it is expected to maximize the agent's interest, *on the condition that* it does not conflict with an action that is ruled to be rational by virtue of principles (1) and (7).

To prepare for the concluding remarks, let me present the ultimate conception of rationality, which I have just reached, as consisting of the following principles:

I. A disposition is rational for *p* to have if and only if it is expected to maximize *p*'s own interest.
II. If it is rational to be disposed to constrained maximization, then it is rational to act in conformity with the disposition.
III. An action is rational if it is expected to maximize the agent's interest.
IV. In case of conflict, rulings based on principle (II) together with principle (I) take precedence over those based on principle (III).

IV

I am now in a position to look back to assess how far Gauthier should have come away from the naive understanding of the maximizing conception of rationality in the attempt to justify the rationality of following the commands of morality. Let me quote Gauthier:

> But Hobbes . . . must revise his conception of rationality, breaking the direct connection between reason and benefit with which he began to reply. Hobbes needs to say that *it is rational to perform one's covenant* even when performance is not directly to one's benefit, *provided that it is to one's benefit to be disposed to perform.* But this he never says. And as long as the Foole is allowed to relate reason directly to *benefit in performance,* rather than *benefit in the dispositions to perform,* he can escape the refutation (MA, 162; my emphasis).

This more than adequately makes my point: when one introduces the principle of disposition maximizing while rejecting the principle of action maximizing, one should be breaking "the direct connection between reason and interest." I made a further point that may have also been implicit in Gauthier's discussions: rejection of the principle of action maximizing in its full scope requires the introduction of the principle governing the connection between the rationality of dispositions and the rationality of actions as an independent principle; the latter can not be a consequence of the core of the maximizing conception.

Suppose now that adding principle (1) is acceptable to Gauthier. Then the previous point about the connection between reason and individual interest needs an elucidation; for it is not merely that the direct connection from individual interest to rationality is not available in an adequate conception of rationality. Rather, the correct picture may be that a constraint to "the direct pursuit of individual interest" is found to be embedded in the concept of rationality itself, however tenuous its force may appear before serious deliberation. The constraint lies in restricting the application of the (weakened) action-maximizing principle (III) in terms of the connecting principle (II) and, perhaps more explicitly, the precedence principle (IV).

If this is a correct analysis, I may come to another elucidation of the nature of Gauthier's moral contractarianism. Again, let me begin by quoting some passages in MA, chapter 1, where Gauthier presents an overview of his theory:

> Morals by agreement begin from an initial presumption against morality, as a constraint on each person's pursuit of her own interest (8f).
> Let us suppose it agreed that there is a connection between reason and interest (8).
> Let it be further agreed that in so far as the interests of others are not affected, a person acts rationally if and only if she seeks her greatest interest or benefit (6f).

These statements indicate that Gauthier believes that a constraint on the direct pursuit of individual interest, *which is absent in the concept of rationality in itself,* can be derived when the direct pursuits need to be compromised in view of *the complexity of the situation in which we are put.* I grant that I might be wrong in diagnosing Gauthier's own view of his project; for instance, the third passage might have been introduced just to explain the action-maximizing conception of rationality which Gauthier rejects in due course. Still, it may be worth noting that the project as I possibly mistakenly understand it does not necessarily conform to what is actually going on in his contractarian justification of morality. What the contractarian moral theory shows may be that what initially does not look like much of a constraint turns out to be substantial and powerful when applied to so complex a situation as we are put in, a situation in which identity and conflict of interests are hard to sort out. If this understanding is correct, I doubt it can be significantly claimed that "[m]orals by agreement begin from an initial presumption against morality" (MA, 8f).

BYEONG-UK YI

University of California/Los Angeles

NOTES AND NEWS

The Editors wish to remind our correspondents that notices of scheduled meetings and symposia are of value to our readers only if they appear early enough to make attendance possible. Copy for *Notes and News* must be planned and processed in this office in time for it to go to the printer at least three months before publication date, the 1st of each month. Accordingly, an announcement of a meeting to be held, say, in April, should be scheduled for our January issue (or sooner) and should be received in this office by September. The Editors appreciate notices that come to us ready for the printer, requiring little editorial work, in JOURNAL style (see these columns), with no imperatives, no prices, no telephone numbers, no unnecessary capitals; and reasonably terse since space is limited. *Notes and News* is included in the JOURNAL when space is available, and on a first come, first served basis; so we cannot guarantee publication of an announcement even if it is received with all the above conditions met.

Analysis is pleased to announce its second Essay Competition. Papers may be on any topic broadly in keeping with the traditional concerns of the journal. An upper limit of 4000 words, including footnotes but excluding any end list of references, will be strictly applied. Entries (two copies, prepared for blind refereeing, accompanied by a word count, and clearly marked "Competition") must arrive no later than March 31, 1993. Authors should be aged 30 or under on that date; but entrants with interrupted academic careers, who believe that there are grounds for relaxing the age rule in their case, may consult the Editor. The winning essays, with a selection of other entries, will be published together as a supplement to the October 1993 issue of *Analysis*. Entries should be sent to *Analysis*, Peter Smith, Editor, Phil. Dept., Univ. of Sheffield, Sheffield, S10 2TN, U.K.

Garland Press is pleased to announce that work has begun on *The Freud Encyclopedia: Theory, Therapy, and Culture*. The Editors of the volume will be Edward Erwin, Michael Moskowitz, and James Walkup. Among the advisors are Adolf Grünbaum, Philip Holzman, and Theodore Shapiro. The Editors' goal is to present unbiased coverage, in alphabetical entries solicited from scholars around the world, of all aspects of Freud's life, theories, and influence. Publication is scheduled for spring 1994. Inquiries should be directed to Prof. Edward Erwin, Editor, Phil. Dept., Univ. of Miami, Coral Gables, FL 33124.

Garland Press is also pleased to announce that work has begun on *The Encyclopedia of Aesthetics*. The Editor will be Michael Kelly of *The Journal of Philosophy*, Columbia University. Among the advisors are Arthur C. Danto, Mary Mothersill, Richard Kuhns, Marx Wartofsky, Joseph Margolis. The Editor's goal is to cover key philosophical topics in the history of aesthetics, including the most recent debates. Publication is expected in 1995. Inquiries should be directed to Michael Kelly, Phil. Dept., 708 Philosophy Hall, Columbia Univ., New York, NY 10027.

Harvard University, The Program in Ethics and the Professions, invites applications for Fellowships in Ethics. Six Fellowships will be awarded in 1993–4 to outstanding teachers and scholars who wish to develop their competence to address ethical issues in faculties of business, government, law, medicine, public policy, and social science. Fellows will participate in the weekly seminar of the Program, attend courses in one of the professional schools or in the Graduate School of Arts and Sciences, and conduct their own research on ethics. The Fellowship usually extends from September through July. Applicants usually hold a postgraduate degree in business, government/public policy, law, or medicine, or a doctorate in philosophy, political theory, or theology. Application deadline is January 6, 1993. For further information, contact Jean McVeigh or Helen Hawkins, The Program in Ethics and the Professions, Harvard Univ., 79 Kennedy St., Cambridge, MA 02138.

The Hoover Institution, with support from the National Endowment for the Humanities, has recently completed a project for the preservation microfilming of the Sir Karl Popper Papers, a large personal archive of over 340,000 documents spanning the whole of Popper's career. Acquired by the Institution in 1986, these materials include holograph and typescript drafts of Popper's published and unpublished writings; outlines and transcripts of speeches and lectures; notes assembled for writing projects (many of which are uncompleted); and letters directly related to publications, projected publications, and speeches. Printed copies of Popper's published articles are included in many cases; printed books are not included. The speeches and writings series is arranged chronologically by date of composition or publication, beginning with the original typescript of Popper's 1928 doctoral dissertation and ending with an unpublished 1986 holograph on probability theory. Also important is the correspondence series with hundreds of physicists, biologists, mathematical logicians, philosophers, social scientists, and classicists. For additional information, please contact the Hoover Institution Archives, Stanford, CA 94305-6010.

The Stanford Humanities Center will offer up to six external fellowships for 1993–4 in the following categories: (1) senior fellowships for well-established scholars; (2) junior fellowships for scholars who at the time of application are at least three years beyond receipt of the Ph.D. and normally no more than ten. For 1992–3, junior fellows were offered stipends of up to $20,000 and senior fellows stipends of up to $30,000. We anticipate offering stipends in a comparable range for 1993–4. The deadline for applications is November 16, 1992. Application materials and further information may be obtained by writing to the Fellowship Program, Stanford Humanities Center, Mariposa House, Stanford Univ., Stanford CA 94305-8630.

The North American Association for the Study of Jean-Jacques Rousseau invites papers for Rousseau and Criticism/Rousseau et la Critique, an interdisciplinary conference to be held at Trent University, Ontario, Canada, May 27–30, 1993. We especially welcome papers on Rousseau and Romanticism, Rousseau and contemporary feminism, and Rousseau's influence on contemporary literary theories. One- or two-page abstracts should be sent by December 1 to Jim MacAdam, Phil. Dept., Champlain College, Trent Univ., Peterborough, Ontario, Canada K9J 7B8.

The Department of Philosophy, Logic and Scientific Method at the London School of Economics and Political Science has established a Lakatos Scholarship for Postgraduate Study in Philosophy in memory of Professor Imre Lakatos who taught in the Department up to the time of his death in 1974. Candidates for the two-year taught M.Phil. Degree in Philosophy, and for the M.Phil/Ph.D. Philosophy research degree are eligible to apply. The Scholarship is tenable for two years, although the second year is conditional on satisfactory academic progress in the first year. Applications for the Scholarship for students entering the 1993/4 session should be submitted to: the Secretary, Phil. Dept., Logic & Scientific Method, The London School of Economics, Houghton St., London WC2 2AE. Closing date: January 31, 1993. Applications for admission to the above degrees should be made to the graduate school in the usual way.

The international quarterly *Philosophical Inquiry* on the completion of its fifteen years of continuous publication has been reorganized by introducing new sections of philosophical interest, increasing its pages and changing its cover. It also plans to include a series of special issues on general philosophical themes as well as special monographs. Scholars and others concerned may submit their papers for the following special issues: Presocratics, Sophists, Socrates, Plato, Aristotle, Stoics, Skeptics, Epicureans, Plotinus, Modern scepticism, Causality, Art and Technology, Aesthetic Criteria, Philosophy of Language, and Chinese Philosophy. Dr. St. Virbidakis, Aristotelian University of Thessaloniki, will function as Assistant Editor. Editor remains Professor D. Z. Andriopoulos, Phil. Dept., Aristotelian Univ. of Thessaloniki, Greece.

Loyola University of Chicago and Loyola University Press have scheduled publication of Volumes IV through VII in their joint *Values and Ethics* Series. For 1992–3 their publication schedule includes works on: the narrative theology of St. Teresa of Avila; abortion based on prenatal diagnosis; antinomies of social theory; and the burdens of distributive justice. Interested authors in humanities, history, philosophy, theology, law, education, political science, sociology, medicine and the helping professions, etc. should send letters of inquiry, CV, an outline of the manuscript, Table of Contents, etc. to: Dr. Roseanne Perez-Woods, Chairperson, Editorial Board, Values and Ethics Series, Loyola Univ. of Chicago Medical Center, 2160 South First Avenue, Bldg. 131-N, Maywood, IL 60153.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 10
OCTOBER 1992

page
499 *Belief Ascription* Stephen Schiffer
522 *In Defense of the Ramsey Test* Sven Ove Hansson

541 NEW BOOKS
549 NEW BOOKS: ANTHOLOGIES

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

1993

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$65.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$10.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U S Copyright Law. This consent does not extend to any other kinds of copying More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY
VOLUME LXXXIX, NO. 10, OCTOBER 1992

BELIEF ASCRIPTION*

I shall do four things in this paper. First, I shall propose a certain theory of the semantics of belief ascriptions as being the best theory of their semantics relative to a certain assumption. Second, I shall raise three problems for this best theory. Third, I shall make what I hope is an interesting connection between the main issue addressed and the vexing question about the form that a meaning theory for a particular language must take. And, fourth, I shall close with a word about the theoretical situation thus determined.

The assumption in question is the widely accepted one that each natural language has a correct compositional truth theory. A compositional truth theory for a particular language *L* is a finitely axiomatizable theory of *L* that issues in a theorem of the form

> An utterance of '*S*' is true in *L* iff such and such.

for each of the infinitely many truth-evaluable sentences of *L*. For example, a truth theory for English might contain a theorem like

> An utterance of 'She wrote it' is true in English iff the female referred to by the utterance of 'she' wrote the thing referred to by the utterance of 'it'.

If we assume that English has a correct compositional truth theory, then we may ask how a particular class of sentences is to be accommodated in such a theory. This is to ask about the kinds of

---

* I am much indebted to the participants in the seminar I gave on these topics in 1991 at the CUNY Graduate Center. Earlier versions of this paper were read in spring 1991 at the University of Wisconsin/Madison, the University of Chicago, and the SOFIA conference in Salamanca, Spain. I benefited considerably from the discussions at these talks, and especially from the comments of Ned Block, James Higginbotham, Richard Jeffrey, Jaegwon Kim, Ernest LePore, Richard Mendelsohn, and Ernest Sosa.

0022-362X/92/8910/499–521

semantic values that would have to be assigned to the components of those sentences and the kinds of compositional rules that would need to apply to those values in order to crank out a truth condition for each sentence of the class. Roughly speaking, this question of accommodation within a compositional truth theory is what philosophers are concerned with when they ask about the *logical form* of a class of sentences. And this is my cue to restate the first of the things I shall do—namely, to offer what I take to be the best account of the logical form of belief ascriptions. In other words, I shall try to say how belief ascriptions would have to be accommodated in a compositional truth theory on the assumption that English has one.

I. THE HIDDEN-INDEXICAL THEORY BALDLY STATED

If English has a compositional truth theory, how should belief sentences be accommodated in it? What, that is to ask, is the best theory of the logical form of belief ascriptions on the assumption that they have a logical form? The theory I shall propose in answer to this question is one I shall call the ***hidden-indexical theory***. It is a theory whose essential idea must have occurred to almost anyone who has thought seriously about the semantics of belief sentences. I first discussed a version of it in "Naming and Knowing";[1] I discuss it briefly in the form I shall presently give it in "The 'Fido'-Fido Theory of Belief";[2] and a version of the theory is nicely presented and defended in Mark Crimmins's and John Perry's "The Prince and the Phone Booth: Reporting Puzzling Beliefs."[3]

The theory, in my favored version, begins with the claim that believing—the relation expressed by 'believes' in a sentence of the form 'x believes that S'—is a three-place relation, $B(x, p, m)$, holding among a believer x, a certain sort of *structured proposition* p, and a *mode of presentation* m under which x believes p. The idea is that x may believe p under one mode of presentation m, disbelieve p under a second mode of presentation m', and suspend judgment altogether under a third mode of presentation m''. Actually, I should have said that believing is a four-place relation, where the fourth place is a time at which x believes p under m, but I shall suppress the temporal reference for simplicity. Let me now explain the notions of a structured proposition and a mode of presentation before moving on to the essence of the theory, its account of the semantics of belief reports.

The sort of proposition I have in mind is a structured entity whose basic components always include at least one property or relation, and may include any sort of contingently existing object. The simplest example of such a structured proposition would be a so-called *singular proposition*, such as the ordered pair ⟨Fido, doghood⟩. This proposition is true just in case Fido instantiates doghood, the property of being a dog; that is to say, just in case Fido is a dog. Interesting technical questions arise about the representation of more complex propositions, but singular propositions will serve present purposes well enough. All that really matters to the hidden-indexical theory is that the propositions we believe are structured entities composed out of the objects and properties we are pretheoretically taken to be talking about when we express our beliefs. I shall call such propositions *Kaplan propositions,* in honor of David Kaplan, who did so much to popularize them.[4]

---

[1] *Midwest Studies in Philosophy*, II (1977): 28–41.

[2] *Philosophical Perspectives*, I (1987): 455–80.

[3] This JOURNAL, LXXXVI, 12 (December 1989): 685–711.

This brings us to modes of presentation. A mode of presentation of a *proposition* is determined by (a) the modes of presentation of the objects and properties contained in the proposition and (b) the place of those things in the proposition. As a first approximation, you might think of a mode of presentation of a proposition as an *n*-tuple of modes of presentation of the objects and properties contained in the proposition.

The expression 'mode of presentation', as it was introduced by Gottlob Frege and as it is used today, is a technical notion, and to understand it is to appreciate the need for its introduction. Two little stories should help remind us of that need.[5]

The first is the infamous case of the morning dog and the evening dog. There is a certain dog who begs at Ralph's door every morning. Ralph feeds this dog, whom he has named 'Fido', and has grown attached to it. Ralph believes that Fido is male. There is also a certain dog who begs at Ralph's door every evening. Ralph feeds this dog, too, whom he has named 'Fi Fi', and has also grown quite attached to it. Ralph believes that Fi Fi is female, and he thinks that Fido and Fi Fi would make a really cute couple. Unbeknown to Ralph, Fido *is* Fi Fi.

---

[4] Let a *singular proposition* be any ordered pair $\langle\langle x_1, \ldots, x_n\rangle, \Phi_n\rangle$, where $\langle x_1, \ldots, x_n\rangle$ is an *n*-ary sequence of items and Φ^n is an *n*-ary property (brackets are customarily dropped for one-membered sequences). Such a proposition is true just in case its *n*-ary sequence instantiates its *n*-ary property. One way to go with Kaplan propositions is to treat them all as singular propositions Thus, the proposition *that Fido is a dog and Pieface is a cat* might be

⟨⟨⟨Fido, doghood⟩, ⟨Pieface, cathood⟩⟩, CONJ⟩

where CONJ is that relation between two propositions that obtains iff both are true, and the proposition that some dogs bark might to a first approximation be represented as

⟨the property of being a dog that barks, SOME⟩

where SOME is that property that a property has iff it is instantiated And so on

[5] The next couple of paragraphs derive from my "The Mode-of-Presentation Problem," in C. Anderson and J. Owens, eds , *Propositional Attitudes* (Stanford: CSLI, 1990).

The second story is that of the spurious natural kind shmoghood. Ralph came upon a race of creatures that he thought comprised a previously unencountered biological species, and he introduced the word 'shmog' to designate members of that species. "A thing shall be called a 'shmog'," Ralph said, "just in case it belongs to the species of these creatures." Unbeknown to him, however, shmoghood is doghood; Ralph had stumbled not upon a new species but upon a new race of dogs, and thus the property that 'shmog' has been introduced as standing for is none other than doghood.

Now, when Ralph says "Fido is male but Fi Fi is female," he is not being irrational in holding the belief which that utterance expresses. But he could not rationally hold the belief that would be expressed by an utterance of 'Fido is and is not male'. Likewise, Ralph is not irrational in holding the two beliefs expressed by a morning utterance of 'This dog is male' and an evening utterance of 'This dog is female', even though the same dog is referred to in both utterances. But Ralph could not rationally hold the belief that would be expressed by a sincere utterance of 'This dog is and isn't male'.

Similarly, when Ralph says "Fido is a dog, not a shmog," and when he says "No dogs are shmogs," he is not being irrational in holding the beliefs those utterances express. But he would be irrational were he to hold the belief that would be expressed by an utterance of 'No dogs are dogs'.

Appeal to the notion of a mode of presentation is designed precisely to accommodate these sorts of data. The intuitive idea, definitive of the notion of a mode of presentation, may be called *Frege's constraint*. Stated informally and in a way that prescinds from the details of any particular theorist's account of belief or of belief-ascribing sentences, Frege's constraint has two parts. First it says that a rational person x may both believe and disbelieve that a certain thing or property y is such and such only if there are distinct modes of presentation m and m' such that x believes y to be such and such under m and disbelieves it to be such and such under m'. Then it says that there are distinct modes of presentation m and m' such that rational person x believes y to be such and such under m and disbelieves y to be such and such under m' only if x fails to realize that m and m' are modes of presentation of one and the same thing. In other words, you cannot rationally believe and disbelieve something under one and the same mode of presentation, or under modes of presentation that you realize are modes of presentation of the same thing. The notion of a mode of presentation is *functionally defined* by Frege's constraint in that something is a mode of presentation if it plays the role defined by Frege's constraint, and nothing can be a mode of presentation unless it plays that role. *To ask what modes of presentation are is just to ask what things play that role.* Presently I shall raise the question of what modes of presentation might be, but the hidden-indexical theory I shall now describe is so far neutral on the outcome of that question.

As applied to the paradigmatic example

[1] Ralph believes that Fido is a dog.

the hidden-indexical theory says that the logical form of an utterance of this sentence may be represented as

[2] $(\exists \boldsymbol{m}) (\Phi^* \boldsymbol{m} \ \& \ \text{B}(\text{Ralph}, \langle \text{Fido, doghood} \rangle, \boldsymbol{m}))$

where Φ^* is an implicitly referred to and contextually determined type of mode of presentation. By a type of mode of presentation I mean merely a property of modes of presentation; Φ^*, for example, might be that property that a propositional mode of presentation has when and only when it requires thinking of Fido as being the dog who appears in the morning and requires thinking of doghood as a property shared by such-and-such similar-looking creatures. The reference to a type of mode of presentation is implicit in that, although the sentence requires the speaker to be referring to a type of mode of presentation whenever the sentence is uttered, there is no word in [1] which refers to that type (whence the 'hidden' in 'hidden-indexical theory'). The reference to the type of mode of presentation is "contextually determined" in that different types may be referred to on different occasions of utterance (whence the 'indexical' in 'hidden-indexical theory').

As we shall see when I stop baldly stating the theory and begin defending it, the reference is to a type of mode of presentation rather than to some particular mode of presentation because we need not be in a position to refer to an actual mode of presentation under which someone believes a proposition. But in a limiting case reference may be to an actual mode of presentation, for the "type" referred to may be one that uniquely picks out a particular mode of presentation, or it may simply be the property of being identical to such-and-such mode of presentation. So at one end of the spectrum, the type referred to may constitute a reference to a unique mode of presentation; at the other end, the type might be vacuous, as when the speaker's import in uttering [1] is merely that Ralph believes the proposition that Fido is a dog under some mode of presentation or other. In between, we have reference to substantial types that do not determine unique modes of presentation.

So the representation of [1] as [2] tells us quite a bit. It tells us that a correct compositional truth theory for English must construe 'believes' as a three-place relational predicate holding among believers, Kaplan propositions, and modes of presentation of those propositions. It tells us that [1]'s 'that'-clause, 'that Fido is a dog', is a referential singular term whose referent is the singular proposition ⟨Fido, doghood⟩, and it tells us that the references of 'Fido' and 'dog' in that 'that'-clause are Fido and doghood, respectively. It tells us that an utterance of the sentence requires reference to a type of mode of presentation, so that an utterance of [1] is true just in case [2], where Φ^* is the mode-of-presentation type referred to in the utterance. And it tells us that this contextually determined reference to a type of mode of presentation is by a "hidden indexical" in that there is no actual indexical in [1] which carries this reference. In this respect, belief ascriptions are like other sentences containing hidden indexicals. For example, in uttering 'It's raining', the speaker must be referring to some place at which it is raining (typically, this is the speaker's location, but it need not be: a speaker in Chicago may reply 'It's raining' when asked about the weather in New York).

II REMARKS IN SUPPORT OF THE HIDDEN-INDEXICAL THEORY

According to the hidden-indexical theory, an utterance of the sentence '*A* believes that *S*' is true just in case the utterer is referring to a type of mode of presentation Φ and the referent of '*A*' believes the Kaplan proposition referred to by 'that *S*' under some mode of presentation of the type Φ. My claim is that this theory provides the best account of how belief ascriptions must be accommodated in a compositional truth theory for English if English has one. A full-scale defense of this claim would outrun a feasible number of pages, but I would like to give *some* sense of what supports the theory.

One very important thing the hidden-indexical theory tells us is that the 'that'-clause in a paradigmatic belief ascription such as [1] is a *referential singular term*. This at least has the *appearance* of being right: [1] does indeed seem to be telling us one of the things Ralph believes—to wit, that Fido is a dog. On the face of it, [1] says that Ralph stands in the belief relation to the referent of the singular term 'that Fido is a dog'. This appearance that [1]'s 'that'-clause, 'that Fido is a dog', is a referential singular term is sustained by the evident validity of arguments such as

Ralph believes that Fido is a dog, and so does Thelma.
So, there is something that they both believe (viz., that Fido is a dog).

Ralph believes everything that Thelma says.
Thelma says that Fido is a dog.
So, Ralph believes that Fido is a dog.

Ralph believes that Fido is a dog.
That Fido is a dog is impossible.
So, Ralph believes something that is impossible (viz., that Fido is a dog).

For how else are we to account for the validity of these inferences other than on the assumption that 'that'-clauses are referential singular terms? I think that the rhetorical force of this question is correct: the best explanation of the way in which these arguments may be valid does presuppose that 'that'-clauses are referential singular terms.[6]

It is obvious that we may give simple and straightforward explanations of the validity in question via hypotheses that entail that 'that'-clauses are referential singular terms. The question is whether validity can be accounted for without that assumption. Perhaps, as I myself argued in *Remnants of Meaning*,[7] the arguments are valid only when their quantifications are read *nonobjectually*, say, as something akin to substitutional quantification. This would entail that 'that'-clauses were *not* referential singular terms. Or perhaps, as Graeme Forbes[8] has argued, we can use some very fancy footwork to account for the validity even though [1]'s 'that'-clause disappears on analysis, leaving us with an existential generalization along the lines of

(∃***m***) (∃***m'***) (***m*** is a mode of presentation of Fido & ***m'*** is a mode of presentation of doghood & B(Ralph, the proposition that ***m*** has ***m'***))

Or perhaps we could salvage enough validity using Donald Davidson's "paratactic" theory of propositional-attitude ascriptions, wherein the only singular term referring to what Ralph believes is the demonstrative 'that', as it occurs in

---

[6] I say 'the way in which these arguments *may* be valid' because of the alleged hidden-indexical nature of belief ascriptions: qua inference *type*,

Ralph believes that Fido is a dog.
So, Ralph believes that Fido is a dog.

is no more valid than

It's raining.
So, it's raining.

Tokens of these inference types are valid only when premise and conclusion make the same implicit reference.

[7] Cambridge: MIT, 1987.

[8] "The Indispensability of *Sinn*," *The Philosophical Review*, xcix, 4 (1990): 535-63

> Ralph believes that. Fido is a dog.

when its reference is the utterance that follows it of 'Fido is a dog'.

Perhaps; but I doubt it. I have argued elsewhere[9] against Davidson's proposal, and if it were not for present constraints I would offer to explain why neither Forbes's line nor my former self's nonobjectual-quantification line can be made to succeed.[10] But there are the constraints; so I must ask you provisionally to assume that I have justified the assumption that 'that'-clauses are referential singular terms. Then we may move to the question this assumption invites.

If 'that Fido is a dog' is a referential singular term, then we may ask what its reference is. To what, that is, does it refer? This, however, is easy to answer: 'that Fido is a dog' refers to that Fido is a dog; what 'that Fido is a dog' refers to is that Fido is a dog. But what is this thing, *that Fido is a dog,* which is the referent of the 'that'-clause singular term? Well: (i) it is ***abstract,*** in that it has no spatial location. That Fido is a dog is not under the kitchen sink or anywhere else. (ii) It is ***mind- and language-independent*** in that it exists in possible worlds in which there are neither thinkers nor speakers. That Fido is a dog is also language-independent in that, while it may be the content of a sentence of any language, it itself belongs to no language; it is not a linguistic entity. (iii) It has a ***truth condition,*** and it has its truth condition ***essentially,*** in contrast to the contingent way sentences have theirs. It is a contingent fact that 'Fido is a dog' is true iff Fido is a dog; if the conventions of English had been different, it might have had a totally different truth condition or none at all. But it is a necessary truth that that Fido is a dog is true iff Fido is a dog. It follows, then, that the referents of 'that'-clauses are *propositions,* in the philosophical sense of that term: abstract, mind- and language-independent objects that have essentially the truth conditions they have.

---

[9] *Remnants of Meaning,* ch. 5.

[10] Well, I guess I can afford to mention one especially devastating problem for the Forbes line Applied to

> Ralph *said* that Fido is a dog

it yields

> $(\exists \boldsymbol{m})$ $(\exists \boldsymbol{m}')$ ($\boldsymbol{m}$ is a mode of presentation of Fido & $\boldsymbol{m}'$ is a mode of presentation of doghood & S(Ralph, the proposition that $\boldsymbol{m}$ has $\boldsymbol{m}'$))

The trouble with this is that it is extremely unlikely that Ralph, in his utterance of 'Fido is a dog', will have said any such mode-of-presentation-containing proposition If he did mean some such proposition, then there would be a specification of what he said that is other than 'that Fido is a dog' and that refers to a mode-of-presentation-containing proposition But it is clear that there need be no such alternative specification of what he said.

If the hypothesis that 'that'-clauses are referential singular terms is to cohere with the assumption that English has a compositional truth theory, then we must suppose that each 'that'-clause is a semantically complex singular term whose reference is determined by its syntax and the references, or extensions, of its component words. What, we might then ask, are the references of 'Fido' and 'dog' in 'that Fido is a dog', as it occurs in an utterance of [1]?

For all intents and purposes, there are but two possible answers. One, apparently Frege's, is that the terms in a 'that'-clause refer to modes of presentation and the proposition referred to by the entire 'that'-clause is a structured proposition made up of those modes of presentation. With respect to a particular utterance of [1], the Fregean proposal would be that the token of 'Fido' refers to a particular mode of presentation of Fido while the token of 'dog' refers to a particular mode of presentation of doghood, the utterance of [1] thus enjoying the representation

$$B(\text{Ralph}, \langle m_{\mathrm{f}}, m_{\mathrm{d}}\rangle)$$

$\langle m_{\mathrm{f}}, m_{\mathrm{d}}\rangle$ being the mode-of-presentation-containing proposition referred to by the occurrence of 'that Fido is a dog' in the utterance. This will be true, of course, only if m_{f} and m_{d} are ways Ralph has of thinking of Fido and doghood, respectively.

The merit of this proposal is that it allows us to see how the ancient astronomer was able to believe that Hesperus was Hesperus long before she believed that Hesperus was Phosphorus. Its demerit is that it is false. For consider

> [3] Everyone who has ever known her has believed that Madonna was musical.

According to the Fregean proposal, there is a particular mode of presentation m of Madonna and a particular mode of presentation m' of the property of being musical such that the foregoing utterance of [3] is true only if everyone who has ever known Madonna has believed the proposition $\langle m, m'\rangle$. Yet this is surely too strong a requirement on the truth of [3]. It requires that everyone who has ever known Madonna shared a single way of thinking of her and a single way of thinking of the property of being musical, and this is most unlikely given that there may have been people who knew her as a child and then died and that someone like Helen Keller may have been among them. If the Fregean proposal were correct, we could prove that [3] was false just on the basis of the virtual certainty that, whatever modes of presentation turn out to be, there will not be one single mode of presentation of Madonna shared by all who ever knew her. But it seems that we cannot disprove [3] in this way, and this because we would count it as true if each person who has ever known Madonna thought of her as being musical *under some mode of presentation or other.*[11]

Our question is: What are the references of 'Fido' and 'dog' in [1]'s 'that'-clause on the assumption that it refers to a proposition that is true just in case Fido is a dog? The second possible answer is that they are Fido and doghood, respectively, ⟨Fido, doghood⟩ being the referent of 'that Fido is a dog'.[12] Certainly this offers a proposition with the right truth conditions. The prima facie problem with this proposal is that it is apt to suggest—and has been taken by many to suggest—that the logical form of [1] may be represented as

[4] B(Ralph, ⟨Fido, doghood⟩)

And you may be very quickly reminded of the well-known problem with this representation if I momentarily change the example. Surely, Lois Lane does not believe that Clark Kent flies, though she does believe that Superman flies. But this is impossible according to the present proposal, since, as Clark Kent = Superman, the proposition that Superman flies = the proposition that Clark Kent flies. Likewise, reverting now to the ongoing example, Ralph may believe that Fido is a dog without believing that Fido is a shmog, and he may believe that Fido is male without believing that Fi Fi is male. Yet all this is impossible according to the theory in question, since it implies that the proposition that Fido is a dog = the proposition that Fido is a shmog and that the proposition that Fido is male = the proposition that Fi Fi is male. The problem is even manifested in belief ascriptions involving indexical reference. For example, pointing to one mug shot, Thelma might say, 'I believe that he is the culprit', while pointing to another, she might say, 'I neither believe nor disbelieve that this guy is the culprit; I simply can't tell'. Intuitively, both of her utterances are true, even though, as it may happen, the same man is referred to in each utterance; but this, again, is incompatible with the proposal in question.

---

[11] Here is a longer way of showing that 'that'-clauses do not refer to mode-of-presentation-containing propositions. Intuitively, 'believes that Fido is a dog' may be univocally true of people who think of Fido and doghood in radically different ways (you and Helen Keller may be among them) and who do not share any single mode of presentation for either Fido or doghood This shows that 'that Fido is a dog' makes no context-independent reference to a mode-of-presentation-containing proposition At the same time, a speaker may truly say that so-and-so believes that Fido is a dog even though she is not in a position to refer to any particular mode of presentation so-and-so has for either Fido or doghood This shows that the 'that'-clause makes no context-dependent reference to a mode-of-presentation-containing proposition. And if it makes neither a context-independent nor a context-dependent reference to such a proposition, then it makes no reference to one.

[12] Friends of possible-worlds semantics or of counterpart theory will think my rush to singular propositions a bit cavalier. Why not say that the reference of 'Fido' is a function from possible worlds to referents of 'Fido' in those worlds—Fido himself, perhaps, in each world, or variable counterparts of our Fido? First, I think importation of possible worlds is unmotivated, though I cannot argue that here. Second, I think any prima facie plausible theory that used possible worlds and propositions that were constructions based on them would, when the dust settled, simply yield alternative styles of bookkeeping. Third, nothing that I really care about changes if, say, you opt for a version of the hidden-indexical theory wherein 'Fido' in a 'that'-clause refers to that constant function that maps each possible world onto our Fido (but the counterpart theory might be a version of the rejected proposal that the referent is a mode of presentation of Fido).

Nathan Salmon[13] has tried to defend the extreme 'Fido'-Fido theory of belief that would represent [1] as [4] by trying to explain away the patina of counterintuitiveness adhering to the theory's consequences. But I think he has not succeeded, and in "The 'Fido'-Fido Theory of Belief" I try to show why.[14] We really do need a theory that allows us to see how an utterance of

> Lois believes that Superman flies but doesn't believe that Clark Kent flies.

can be literally true.

Here, then, is our situation. We are motivated to see [1]'s 'that'-clause as a referential singular term whose reference is the proposition that Fido is a dog. Since we are assuming that English has a compositional truth theory, we must hold the reference of this 'that'-clause to be determined by its syntax and the semantic values the words in the 'that'-clause have in it. Those semantic values—those references—cannot be modes of presentation, and the only viable option is that they are Fido, for 'Fido', and doghood, for 'dog'. At the same time, we cannot accept the representation of [1] as [4], for we also want to allow that Ralph does not believe that Fido is a shmog.

Enter now on its white charger the hidden-indexical theory, which satisfies all desiderata. This theory, you will recall, represents the logical form of an utterance of [1] as [2], where Φ^* is an implicitly referred to and contextually determined type of mode of presentation. Thus, 'that Fido is a dog' in [1] is a referential singular term whose referent is the singular proposition ⟨Fido, doghood⟩, and thus the referents of 'Fido' and 'dog' in that 'that'-clause are Fido and doghood, respectively. At the same time, we can see how an utterance of

---

[13] *Frege's Puzzle* (Cambridge: MIT, 1986).

[14] See also Salmon's reply, "Illogical Belief," *Philosophical Perspectives*, III (1989) 243–85.

[5] Ralph does not believe that Fido is a shmog.

can also be true, even though, as shmoghood = doghood, 'shmog' in [5] has the same reference as 'dog' in [1], so that the 'that'-clause in each sentence refers to one and the same singular proposition, to wit, ⟨Fido, doghood (i.e., shmoghood)⟩. That the utterances of [1] and [5] can both be true is due, of course, to the hidden-indexical feature: in the normal case, the type of mode of presentation referred to in an utterance of [1] will be different from that referred to in an utterance of [5]. Likewise in the other examples discussed, and likewise as regards the London/Londres and Paderewski examples that puzzled Saul Kripke.[15] These are no puzzles for the hidden-indexical theory.

III. THREE PROBLEMS FOR THE HIDDEN-INDEXICAL THEORY

So much for what recommends the hidden-indexical theory. I shall now raise three problems for the theory.

The candidate problem. This problem arises for *any* theory that appeals to modes of presentation. It is a problem I have discussed at length elsewhere,[16] and I will say only enough about it now to indicate its nature. In a nutshell, the candidate problem is that there is some question whether there is a plausible candidate for what modes of presentation are.

The notion of a mode of presentation is, we saw, a functionally defined technical notion: a mode of presentation is whatever plays the mode-of-presentation role defined by Frege's constraint. So we may ask what things play that role. This is like asking what genes are, the notion of a gene being functionally defined as whatever is responsible for the transmission of hereditary characteristics. There the answer turned out to be segments of DNA molecules. What, though, is the answer to the mode-of-presentation question? There is no dearth of proposed answers:

---

[15] "A Puzzle about Belief," in A. Margalit, ed., *Meaning and Use* (Dordrecht: Reidel, 1979).

[16] "The Basis of Reference," *Erkenntnis*, XIII, 1 (1978): 171–206; *Remnants of Meaning*, esp. ch 3; and "The Mode-of-Presentation Problem."

Modes of presentation are *individual concepts,* uniqueness properties of the form *the property of being the unique instantiator of such-and-such property.*

Modes of presentation are simply *general properties* that make no pretense of applying uniquely to the things of which they are modes of presentation.

Modes of presentation are *percept tokens,* actual instances of sensory experience.

Modes of presentation are "*stereotypes,*" perhaps what cognitive psychologists call *prototypes.*

Modes of presentation are "*characters,*" functions from contexts of the utterance of an expression to the expression's contents in those contexts.

Modes of presentation are *public language expressions.*

Modes of presentation are *Mentalese expressions,* formulae in the neural language of thought.

Modes of presentation are *functional roles,* perhaps conceptual roles of Mentalese expressions.

Modes of presentation are *causal chains* linking Mentalese names and predicates to the objects and properties for which they stand.

And finally there is the *no-theory theory of modes of presentation,* which rejects what I have elsewhere called the *intrinsic-description constraint.*[17]

This constraint holds that if a thing is a mode of presentation—if, that is, it plays the mode-of-presentation role—then it must be intrinsically identifiable in a way that does not describe it as a mode of presentation or as a possible mode of presentation. If a thing is a mode of presentation, then it must be intrinsically identifiable as some other kind of thing. Rejecting this constraint, the no-theory theorist of modes of presentation holds that modes of presentation are modes of presentation and there is an end to it. You simply cannot say what they are in any other terms.

But I think the intrinsic-description constraint *is* well-motivated, and I think there are compelling objections against each of the foregoing candidates but one. The exception is the view that modes of presentation are functional roles, perhaps conceptual roles of Mentalese expressions. This view of what modes of presentation are has intrinsic problems and it does not cohere with every theory of the logical form of belief ascription; but its problems do not *refute* it and it does cohere with the hidden-indexical theory.[18] In any case, the hidden-indexical theory entails that believing is a relation among a believer, a proposition, and a mode of presentation of that proposition, which mode of presentation is in turn determined by modes of presentation of the proposition's components. It is hard to see how we can reasonably accept this theory without some well-motivated idea of what these modes of presentation are supposed to be.

---

[17] "The Mode-of-Presentation Problem."

[18] This corrects a flawed argument in "The Mode-of-Presentation Problem" which was intended to show that modes of presentation cannot be conceptual roles of Mentalese expressions. The argument overlooked what must be the in-

The meaning-intention problem. This problem is that one may reasonably doubt that belief ascribers mean what the hidden-indexical theory requires them to mean when they ascribe beliefs. Let me explain.

Floyd's sister calls him in Chicago and asks about the weather. Floyd replies, "It's raining." Here Floyd refers to Chicago, and this by virtue of the fact that in uttering 'It's raining', Floyd *means* that it is raining in Chicago. In other words, Floyd counts as having referred to Chicago because the proposition he meant is about Chicago. Notice that there is no difficulty whatever in ascribing to Floyd the propositional speech act in question: he clearly intended his sister to believe that it was raining *in Chicago,* and he is quite prepared to tell you that this is what he meant, what he implicitly said, and what he intended her to be informed of. Because these things are so clear, Floyd's utterance is a paradigm of implicit reference.

Now for a different case. During a casual conversation about airfare bargains, Flora says,

> [6] Harold believes that TWA is offering a New York-Paris return fare for $318.

According to the hidden-indexical theory, there is a property Φ of modes of presentation of the proposition that TWA is offering a New York-Paris return fare for $318 such that Flora referred to Φ in her utterance. If this is true, then, presumably, it is because Flora, in producing her utterance, meant some proposition about Φ in just

---

tended generalization of the hidden-indexical theory for belief ascriptions with embedded 'that'-clauses—namely, that a belief ascription with n 'that'-clauses embedded in the primary 'that'-clause must be making $n + 1$ implicit references to types of modes of presentation. Thus, the logical form of an utterance of

> Floyd believes that Lois believes that Superman flies.

is

$$(\exists m)\ (\Phi m \ \&\ B(\text{Floyd},\ [(\exists m')\ (\Psi' m' \ \&\ B(\text{Lois},\ [\text{Superman flies}],\ m'))],\ m))$$

where Φ and Ψ' are implicitly referred to and contextually determined types of modes of presentation (and where '[. . .]' is a singular-term forming operator with the force of 'the proposition that . . .'). If this inevitable complexity weakens the plausibility of the hidden-indexical theory, then so be it.

the way that Floyd, in his utterance, meant some proposition about Chicago. But, we shall see, it is doubtful that Flora meant any such thing.

We had no trouble saying what Floyd meant in uttering 'It's raining': he meant that it was raining in Chicago. But what that implies a reference to a type of mode of presentation does Flora mean in uttering [6]? It must surely be

> [7] that there is something that both has Φ and is such that Harold believes the proposition *that TWA is offering a New York-Paris return fare for $318* under it

for some particular property Φ of modes of presentation of the proposition that TWA is offering a New York-Paris return fare for $318. But—and this is where the trouble begins—neither Flora nor her audience is aware of her meaning any such thing. Floyd is consciously aware of both the *form* of the proposition he meant (he is aware that he meant that it is raining *in* Chicago) and the implicitly referred to thing the proposition is about (he is aware that he meant that it is raining in *Chicago*). But it is doubtful that the nonphilosopher Flora has conscious access to the form of [7], and it is especially doubtful that she has conscious awareness of referring to any mode-of-presentation property. That is to say, if she did mean a proposition of form [7], then she has no conscious awareness of what property Φ that proposition is about, and no conscious awareness, therefore, of what she meant in uttering [6].

Neither we nor Flora can say what mode-of-presentation property was referred to in her utterance of [6]. Just try to say it. The referred-to property, if it exists, would be given by a completion of the form

> The property of being an *m* such that to believe the proposition *that TWA is offering a New York-Paris return fare for $318* under *m* requires thinking of TWA, the offering relation, New York, Paris, the US dollar currency, the number 318 [and so on for the other components of the proposition] in . . . ways, respectively.

But the nonphilosopher Flora has no access even to this *form* of specification, and who among us can offer to replace the three dots? Now, of course, it is in principle possible to have knowledge of such a property even though one cannot specify it in the forgoing way. One might, for example, have knowledge of it under some quite extrinsic description. But I submit that it will be obvious on reflection that Flora has no such alternative way of explicitly picking out a property of modes of presentation to which she is implicitly referring in her utterance of [6].[19] There is no sentence σ such that Flora can say 'I meant that σ', where 'that σ' explicitly refers to a proposition of [7]'s form.

Thus, if the hidden-indexical theory is correct, then Flora has no conscious awareness of what she means, or of what she is saying, in uttering [6], and this is a prima facie reason to deny that she means what the theory is committed to saying she means.

There is more than one way a hidden-indexical theorist might respond to this prima facie problem, but let me go directly to what I think is the theorist's best bet. This response is to concede that Flora is not *aware* of meaning a proposition of form [7] under any mode of presentation, but to insist that she nevertheless meant such a proposition. The idea is to appeal to so-called "tacit belief" and "tacit intention." This idea has its origin, I believe, in Noam Chomsky's controversial claim that speakers know a grammar of their language. To this it was protested that in no ordinary sense of 'know' do speakers know any such thing. When the dust settled, many theorists supposed that the best construal of Chomsky's claim was that the correct theory of language processing implies that a speaker has an internal representation of the grammar of her language, where this internal representation is stored not as a *belief*, but as what Stephen Stich was later to call a *subdoxastic state*.[20] Subdoxastic states are representational states that are belief-like in some respects but differ from paradigmatic beliefs in their unavailability to consciousness and the way they are inferentially isolated from beliefs (for example, if you believe that if p, then q, and believe p, you are likely to infer q; but you will not make this inference if the state representing p is subdoxastic). Now, the notion of a tacit propositional attitude can admit of a spectrum of cases, from the Chomsky-required subdoxastic states at one end, to states that may reasonably be claimed to be subsumable under our current propositional-attitude concepts, at the other. What is evidently unifying about the notion of a tacit propositional attitude is that a tacit propositional attitude is unavailable to consciousness and that one cannot be made irrational by conflicts between those of one's thoughts which are available to consciousness and those which are merely tacit. At any rate, the present proposal is that it is thanks to her *tacit* beliefs and intentions in uttering [6] that Flora means a proposition of [7]'s form even though she is not aware of meaning any such thing.

---

[19] Besides, it is quite against the spirit of the hidden-indexical theory that the type of mode of presentation referred to should be one that the speaker knows only under some nonintrinsic mode of presentation. The idea of the theory is that, though one might not know the particular mode of presentation under which so-and-so believes that such-and-such, still one has access to some interesting *description* that the mode of presentation satisfies.

[20] "Beliefs and Subdoxastic States," *Philosophy of Science*, XLV (1978): 499–518.

One trouble with the tacit-intention proposal is that it induces a rather radical *error theory:* not only do ordinary belief ascribers have no conscious knowledge of what they are asserting, they also turn out not to have the conscious thoughts they think they have. Flora clearly *thinks* she has conscious knowledge of what she is asserting in uttering [6]. She is quite prepared to say, "Look, what I am saying, and all that I am saying, is that Harold believes that TWA is offering a New York-Paris return fare for $318." In other words, she thinks she is consciously aware of what she is asserting in uttering [6], but the tacit-intention line implies that here she is in error: the only proposition she asserts in uttering [6]—viz., some proposition of form [7]—is not anything of which she is conscious. What makes this error aspect of the tacit-intention proposal problematic is not merely that it riddles the propositional-attitude ascriptions of ordinary speakers with error; it also forces us to qualify our views about first-person authority in an important way. Flora does not have the privileged access to what she consciously means, intends, and believes in uttering [6] which one might reasonably have supposed to be part of a normal person's functional architecture.

A more serious trouble with the tacit-intention proposal actually constitutes a direct objection to the claim that speakers have the meaning intentions the hidden-indexical theory requires them to have, whether or not those intentions are claimed to be tacit. But the objection is best presented here, in response to the tacit-intention proposal, since it is obvious that speakers do not consciously mean what the hidden-indexical theory entails that they mean. The objection takes hold even if it is conceded that ordinary people do have some sort of indirect awareness of modes of presentation, which is all to the good, as it is not out of the question that one could motivate the idea that Flora tacitly *believes* that Harold believes the proposition about TWA under a certain type of mode of presentation. The point to be pressed, however, is that considerably more is required of tacit *meaning*. I shall explain.

Meaning entails audience-directed intentions, and one cannot mean something without intending to be understood. Part of meaning that such and such is intending one's audience to recognize that that is what one meant, and—a corollary—part of referring to a thing is intending one's audience to recognize that reference. Consequently, the present tacit-intention proposal requires there to be some type of mode of presentation Φ such that Flora tacitly means that Harold believes that TWA . . . under a mode of presentation of type Φ, and tacitly intends her audience to recognize (no doubt tacitly) that she means that Harold believes that TWA . . . under a mode of presentation of type Φ. But as James Higginbotham pointed out to me, this puts the implausibility of Flora's *meaning* a proposition of form [7] on a par with that of a certain way of trying to extend the description theory to "incomplete" definite descriptions. The idea there is that, when a speaker utters a sentence like 'The dog has fleas', then there must be some property Φ such that she means the proposition that the thing that is uniquely a dog and Φ has fleas. What makes this so implausible is that there will typically be a number of potentially completing descriptions that are equally salient in the context (e.g., *the dog that I own, the dog we are both looking at, the spotted dog in the green chair,* and so on). Consequently, it is highly implausible that there will be *one* such description such that the speaker intends it to be understood between her and her audience that she means a proposition containing that very description.[21]

Now, just the same problem arises with the meaning claim required by the hidden-indexical theory. For assume that propositions are believed under modes of presentation. If a proposition is believed under one mode of presentation, then it will typically be believed under many modes of presentation. Further, each of those modes of presentation will instantiate infinitely many *types* of modes of presentation, many of which will be equally salient in the communicative context. This makes it extremely implausible that of all the equally salient types of ways that Harold has of believing the proposition about TWA, Flora should mean—and intend to be taken to mean—a proposition about one definite one of them.

A little thought experiment should make the objection vivid. When considering the hypothesis that a certain expression has a certain semantic property, it is often helpful to introduce an expression that has that semantic property by stipulation. Presently in debate is the hidden-indexical theory's claim that one uttering [6] would mean a proposition of the form [7] displays. So let us introduce 'shmelieves' as a term that satisfies the hidden-indexical theory by stipulation: to utter a sentence of the form

---

[21] See my "Indexicals and the Theory of Reference," *Synthese,* XLIX, 1 (1981):43–100; esp. pp. 77–8.

> *A* shmelieves that *S*.

with its literal meaning requires a speaker to mean

> that there is something that both has Φ and is such that *A* believes the proposition that *S* under it

for some particular contextually determinate property Φ of modes of presentation of the proposition that *S*. The hidden-indexical theory may or may not be true of '*A believes* that *S*', but it is true by stipulation of '*A shmelieves* that *S*'.

So far so good, but let me now try to use this newly minted coin to make an assertion.

> Placido Domingo shmelieves that Luciano Pavarotti will be arriving at Orly airport today at around 4 PM Paris time.

Very well, what assertion did I make? What is the type of mode of presentation Φ such that I meant that Placido Domingo believes the proposition *that Luciano Pavarotti will be arriving at Orly airport today at around 4 PM Paris time* under a mode of presentation that has Φ? Remember, Φ is made up of types of modes of presentation for each of the constituents of the proposition to which my 'that'-clause refers: Luciano Pavarotti, the arrival relation, Orly airport, the number 4, Paris, etc. The question, of course, cannot be answered. For although 'shmelieves' was stipulatively given a perfectly good hidden-indexical semantics, I was not, in uttering the displayed sentence, in a position assertively to use it, and this because I was not in a position to refer to a contextually determinate type of mode of presentation. None is sufficiently salient to enable you, my audience, to identify it as the one I meant, and this notwithstanding the fact that we understand all the concepts involved and everything has been painstakingly raised for us to the level of conscious awareness.

The application to the hidden-indexical theory should be clear. Although I cannot utter the forgoing 'shmelieves' sentence and mean what its literal meaning requires me to mean, there are no such impediments to my uttering 'Placido Domingo *believes* that Luciano Pavarotti will be arriving at Orly airport today at around 4 PM Paris time'. I can perfectly well utter this and hope to have my literal assertion be perfectly well-understood by you. Since 'shmelieves' is by stipulation what 'believes' would be if the hidden-indexical theory were true, this appears to show that the theory is not true. The best account of why neither Flora nor her audience (who understood her perfectly well) is aware of her meaning a proposition of form [7], as the hidden-indexical theory requires her to mean, is not that the requisite meaning intentions have gone tacit; it is that she does not mean any such thing.

The logical-form problem. According to the hidden-indexical theory, 'believes' is a three-place relational predicate. But this might strike one as fishy for the following reason. If 'believes' really were three-place, then it ought to have occurrences where its three-place form is explicit, as the three-placedness of 'gives' is made explicit when we move from 'She gave the house' to 'She gave the house to her husband'. Yet in what analogous way can we make explicit the three-placedness of 'Ralph believes that Fido is a dog'? Perhaps in the way of

[8] Ralph believes that Fido is a dog *in way w/under mode of presentation m.*

There are, however, two problems with this.

First, this is no ordinary-language specification but technical jargon. If one is told that Mary gave the house, one can ask, "To whom did she give the house?" But what is the corresponding ordinary-language question for 'Ralph believes that Fido is a dog'? Ask a nonphilosopher, "In what way, or under what mode of presentation, does Ralph believe this?" and your best answer will be a puzzled look.

Second, [8] does not look like a specification of a three-place relation. It looks more like a two-place relation with *an adverbial qualifier*. In other words, it is to be assimilated to

Ralph kissed her in the most exciting way.

where 'the most exciting way' is merely part of an adverb, rather than to

[9] Mary gave the house to her husband.

where 'her husband' is an argument of the verb 'to give'. Evidence for this assimilation comes from well-known constraints in syntactic theory which distinguish between arguments and adverbs in their ability to extract from 'whether'-clauses.[22] Thus, the argument status of 'her husband' in [9] is revealed in the fact that we can answer 'Her husband' in response to the question

To whom did you wonder whether Mary gave the house?

---

[22] See Chomsky, *Barriers* (Cambridge: MIT, 1986). I owe this reference and the point about 'whether'-clauses to Robert Fiengo.

And the nonargument, adverbial status of 'way *w*/mode of presentation *m*' in [8] is revealed by the fact that it *cannot* be given in answer to the question

> In what way/under what mode of presentation did you wonder whether Ralph believes that Fido is a dog?

IV. THE IMPORTANCE OF THE ISSUE

I have proposed that, if natural languages have compositional truth theories, then the hidden-indexical theory is the correct account of the accommodation of belief ascriptions in them. I have also offered some reason for thinking the hidden-indexical theory is false. If it is both conditionally correct and false, then it follows that natural languages do not have compositional truth theories. This is one reason the issue of the semantics of belief ascriptions is important. There is, however, another reason for the importance of the question of the accommodation of belief ascriptions in a compositional truth theory; it has to do with the bearing of this question on the vexing question about the form that a compositional *meaning* theory for a language must take. I shall explain.

It is commonly assumed that each natural language *L* has a *compositional meaning theory,* this being a finitely axiomatizable theory of *L* whose theorems ascribe to each sentence of *L* its meaning in *L*. Those who make the assumption reason that without it there will be no explaining either the platitude that the meaning of a sentence is determined by its syntax and the meanings of its component words or the fact that we have the ability to understand indefinitely many novel sentences. But while it is generally agreed that every natural language enjoys a compositional meaning theory, no three philosophers can be made to agree on the form such a theory must take. For a start, it is not even clear what form the *theorems* of a meaning theory would have to take. You might at first blush suppose that a compositional meaning theory for *L* would generate theorems of the form

> '*S*' means in *L* that such and such.

but this could hardly be the form of a theorem for sentences like 'He was there' or 'Stop singing, you idiot!' and it is anything but clear that the word 'means' should ever be used in saying what a sentence means. After all, to know the meaning of a sentence is to be able to know what a speaker is saying in uttering the sentence; so it is reasonable to suppose that a compositional meaning theory would have to ascribe to each sentence a feature, knowledge of which would enable us to know what would be said in the utterance of any sentence having that feature. But there is no agreement as to what such a feature would be like, or that the word 'meaning' would need to be used to ascribe it.

Those who assume that each natural language has a compositional *meaning* theory typically assume that each natural language also has a compositional *truth* theory, in the sense already glossed. After all, it is part of the *meaning* of 'Snow is white' that an utterance of it is true iff snow is white; so if it is reasonable to think a compositional theory can assign meanings to sentences, then it is reasonable to think one can assign truth conditions to them, too. What we are now in a position to see is that, if we know how to accommodate belief ascriptions in a compositional truth theory, then we will be able to answer the question about the form that a compositional meaning theory must take. This ought at first to seem a little surprising. Merely knowing how to accommodate 'Fido is a dog' in a truth theory would not per se tell us the first thing about how to accommodate that sentence in a *meaning* theory. So how can knowing how to accommodate *belief ascriptions* in a truth theory tell us how to accommodate *any* sentence in a meaning theory?

I shall forgo a general answer, but we can see straightway how the hidden-indexical theory of belief ascription constrains the nature of a meaning theory. For what applies to belief ascriptions applies, mutatis mutandis, to all propositional-attitude ascriptions involving 'that'-clauses. What holds for 'Ralph believes that Fido is a dog' also holds for 'Ralph says that Fido is a dog'. The theory tells us that the 'that'-clause in a saying ascription refers to a Kaplan proposition. This means that to know what a speaker said in the utterance of a sentence is to know what Kaplan proposition she said. Since to understand a language is to have the ability to understand utterances in it, and since to understand an utterance is to know what the speaker said in it, we can see that, if the hidden-indexical theory is correct, then the main task of a compositional meaning theory for a language will be the construction of a finitely specifiable correlation that relates each sentence of the language to what Kaplan has called a *character*—a function that, in effect, maps each utterance of the sentence onto the Kaplan proposition that is the propositional object of the primary speech act performed in the utterance.[23] In other words, if the hidden-indexical theory of the accommodation of propositional-attitude ascriptions in a truth theory is correct, then so is the "direct reference" approach to a compositional meaning theory, and in this way we can see how the hidden-indexical theory constrains the nature of a meaning theory.

---

[23] Kaplan, "Dthat," in P Cole, ed , *Syntax and Semantics 9: Pragmatics* (New York: Academic, 1978)

It may be worth noting that this point I have been laboring reverses the usual procedure of those who argue for the direct reference treatment of 'that'-clauses implied by the hidden-indexical theory. That procedure begins by assuming a direct reference approach to meaning and then draws the consequence as regards the propositional content of propositional-attitude ascriptions, which consequence entails that 'that'-clauses are referential singular terms whose referents are Kaplan propositions. My procedure has been different, and I think less contentious. I started with no assumptions about "content" or meaning. My question was wholly about the *references* that must be assigned to words in a belief ascription on the assumption that languages had compositional *truth* theories, theories whose theorems are of the extensional paradigm

> 'La neige est blanche' is true in French iff snow is white.

It was then argued that, relative to the assumption, 'that'-clauses had to be construed as referential singular terms whose *referents* are Kaplan propositions. Then, in the final step, this premise about *reference* was used to derive the general point about *meaning*.

V CONCLUSION

If the hidden-indexical theory is correct, then we not only know how to accommodate belief ascriptions in a compositional truth theory for English, we also know the form that a compositional meaning theory for English must take. But I did not argue that the hidden-indexical theory is correct. I merely argued that it is the correct theory relative to the assumption that natural languages have compositional truth theories, and I proceeded to raise problems for the theory. There are but three possible ways out of this impasse: either the problems can be answered, or the hidden-indexical theory is not the best accommodation of belief sentences in a compositional truth theory, or natural languages such as English do not have compositional truth theories. Elsewhere I have argued that the third option is one that deserves to be taken seriously.[24] Others may disagree; but I hope it will be agreed that the hidden-indexical theory and its obstacles present a challenge that cannot be ignored.

STEPHEN SCHIFFER

CUNY Graduate Center

---

[24] *Remnants of Meaning;* "Does Mentalese Have a Compositional Semantics?" in B. Loewer and G. Rey, eds., *Meaning in Mind: Essays on the Work of Jerry Fodor* (Cambridge: Blackwell, 1991); and "Actual-Language Relations," forthcoming in *Philosophical Perspectives*, VII (1993).

IN DEFENSE OF THE RAMSEY TEST[*]

For more than two decades, much of the philosophical debate on conditionals has focused on a proposal by Frank Ramsey[1] for the evaluation of conditional sentences. According to Ramsey, "[i]f two people are arguing 'If *p* will *q*?' and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and are arguing on that basis about *q*" (*ibid.*, p. 247). The proposal was developed into a general principle by Robert Stalnaker.[2] In modern studies of belief dynamics, the Ramsey test is taken to "presume . . . some method of revising states of belief."[3] Thus, 'If *p* then *q*' is taken to be believed if and only if *q* would be believed in the belief state that would result from revising the present belief state by *p*.

Formal studies of the epistemological application of the Ramsey test have gained impetus from the models of belief dynamics that have been developed in particular by Isaac Levi and by C. E. Alchourrón, Peter Gärdenfors, and David Makinson.[4] Gärdenfors, however, has caused considerable commotion in the belief dynamics community by proving that the Ramsey test is incompatible with what seem to be weak and reasonable demands on a system of belief revision.[5]

Section I summarizes Gärdenfors's impossibility theorem and the major escape routes from it. In section II, both informal arguments and a formal result are presented in support of one of these escape routes which leaves the Ramsey test untouched. In section III, a formal account of epistemic conditionals is proposed that complies with the Ramsey test. This account is extended in section IV to iterated conditionals of unlimited complexity.

---

[*] I would like to thank Sten Lindström, Wlodek Rabinowicz, and Krister Segerberg for valuable comments.

[1] *Foundations of Mathematics and Other Logical Essays* (New York: Routledge, 1931).

[2] "A Theory of Conditionals," in Nicholas Rescher, ed., *Studies in Logical Theory, American Philosophical Quarterly Monograph Series, II* (Cambridge: Blackwell, 1968), pp. 98–112, esp. 101–5.

[3] Peter Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (Cambridge: MIT, 1988), p. 148.

[4] Levi, *The Enterprise of Knowledge* (Cambridge: MIT, 1980); Alchourrón, Gärdenfors, and Makinson, "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision," *Journal of Symbolic Logic*, L (1985): 510–30; Gärdenfors, *Knowledge in Flux*

[5] "Belief Revisions and the Ramsey Test for Conditionals," *Philosophical Review*, XCV (1986): 81–93

0022-362X/92/8910/522–40

I THE IMPOSSIBILITY THEOREM

Gärdenfors showed the Ramsey test for conditionals to be incompatible with a set of conditions on an operator * of belief revision (*ibid.*). His proof makes use of the following assumptions:

(L) The language is governed by a logic including classical truth-functional logic, and the consequence relation satisfies the deduction theorem.
(S) Belief states are identified with belief sets, i.e., sets of sentences that are closed under logical consequence.
(0) Belief sets include sentences that contain the conditional connective $>$ as elements.
(1) $p \in K^*p$ (success).
(2) If K and p are consistent, then so is K^*p.
(R) $p > q \in K$ if and only if $q \in K^*p$ (Ramsey test).
(P) If $\neg p \notin K$ and $q \in K$, then $q \in K^*p$ (preservation).
(E) Let K be a belief set. Then $K + p = \{q \mid p \rightarrow q \in K\}$, the expansion of K by p, is a belief set (closure under expansion).
(NT) There are three sentences p_1, p_2, and p_3 and some belief set K such that $p_1 \,\&\, p_2$, $p_1 \,\&\, p_3$, and $p_2 \,\&\, p_3$ are all inconsistent and that $\neg p_1 \notin K$, $\neg p_2 \notin K$, and $\neg p_3 \notin K$ (nontriviality).

From (R) the following follows directly:

(M) For all belief sets K and K' and all sentences p: if $K \subseteq K'$ then $K^*p \subseteq K'^*p$.

Gärdenfors was able to prove that (L), (S), (0), (1), (2), (M), (P), (E), and (NT) are incompatible. From this follows directly that (L), (S), (0), (1), (2), (R), (P), (E), and (NT) are incompatible. (Some of these conditions were not mentioned explicitly by Gärdenfors in his statement of the theorem, but were only included as background conditions. The importance of (E) was first highlighted by Hans Rott.[6])

In Gärdenfors's opinion, we have to give up either (R) or (P), since the others are "rather innocuous background assumptions."[7] His own choice was to give up (R). Thus, according to its author, the impossibility theorem urges us to give up the Ramsey test for epistemic conditionals. Since then, however, several other escape routes have been investigated.

Various ways to weaken (L) have been studied, but the theorem has resisted all attacks on this flank. Thus, Krister Segerberg[8] generalized the impossibility theorem to a large family of nonclassical logics whose consequence relations satisfy monotony (i.e., if X is a subset of Y then $Cn(X)$, the set of logical consequences of X, is a subset of $Cn(Y)$). Makinson[9] has generalized it to important categories of nonmonotonic logics that include classical truth-functional consequence.

---

[6] "Conditionals and Theory Change: Revisions, Expansions, and Additions," *Synthese*, LXXXI (1989): 91–113.

[7] Gärdenfors, "Belief Revisions," p. 86.

[8] "A Note On An Impossibility Theorem of Gärdenfors," *Noûs*, XXIII (1989). 351–4.

Several weakenings of (R) have been investigated by Rott, Gärdenfors, and others.[10] The only weakening that seems to resist (reconstruction of) the impossibility theorem is a version developed by Sten Lindström and Wlodzimierz Rabinowicz,[11] in which revision (*) is not applied to the belief set itself but to maximal consistent extensions of the belief set.

The validity of (0) has been questioned by Levi,[12] who maintains that conditional statements are not "objects of belief which bear truth values and are subject to appraisal with respect to epistemic possibility and probability" (*ibid.*, pp. 49–50). Instead, conditionals are "appraisals of serious or epistemic possibility relative to transformations of the current corpus of knowledge or belief set" (*ibid.*). Therefore, conditional sentences should not be included in belief sets. (See below, section III.)

Rott has argued against (E), i.e., against the assumption that the set of belief sets is closed under logical expansion. In his view "[e]xpansions are not the right method to 'add' new sentences if the underlying language contains conditionals which are interpreted by the Ramsey test."[13]

As will be seen in the following section, I agree with Rott that the Ramsey test can be saved, and that (E) is the culprit. (More precisely, it is the combination of (E), (S), and (0) which I hold to be untenable). My arguments are not the same as Rott's, however, and my formal proposals are different from his (and much more related to proposals by Levi).[14]

II ESCAPING THE IMPOSSIBILITY THEOREM

I shall leave aside the—certainly very important—metaphysical issue whether or not a conditional sentence can be true or false in the same sense as an indicative sentence. For my present purposes it will be sufficient to make the weak, and likely uncontroversial, assumption that belief states can *support* certain conditional sentences.

---

[9] "The Gärdenfors Impossibility Theorem in Non-Monotonic Contexts," *Studia Logica*, XLIX (1990): 1–6.

[10] Rott, "Ifs, Though and Because," *Erkenntnis*, XXV (1986): 345–70; Gardenfors, "Variations on the Ramsey Test: More Triviality Results," *Studia Logica*, XLVI (1987): 319–25, and *Knowledge in Flux*, pp. 160–2.

[11] "Belief Revision, Epistemic Conditionals and the Ramsey Test," forthcoming in *Noûs*.

[12] "Iteration of Conditionals and the Ramsey Test," *Synthese*, LXXVI (1988): 49–81.

[13] Rott, "Conditionals and Theory Change," p. 100.

[14] Levi, "Iteration of Conditionals."

In this section, no assumptions will be made about the structure of belief states. In particular, they may or may not be sets of sentences. From a formal point of view, this means that there is a *support function* s such that for any belief state $\mathcal{K}$, $s(\mathcal{K})$ is the set of sentences (of all categories) that are supported by $\mathcal{K}$. $s(\mathcal{K})$ will include conditional sentences. Depending on the underlying language, it may also include sentences that contain operators of necessity, possibility, knowledge, belief, etc.

Let $\mathcal{L}_0$ be the indicative part of the language, i.e., the part that consists only of sentences of truth-functional or predicate logic. Furthermore, let $\mathcal{L}_1$ be the subset of the language that contains, in addition to the symbols of $\mathcal{L}_0$, only the symbol $>$ for conditionals. Then $\mathcal{L}_0 \cap s(\mathcal{K})$ is the set of indicative sentences that are true in, or supported by, $\mathcal{K}$. (The two locutions are synonymous for indicatives.) $\mathcal{L}_1 \cap s(\mathcal{K})$ contains, in addition, those sentences containing $>$ which are supported by $\mathcal{K}$. (According to some metaphysical views, these sentences, too, are true in $\mathcal{K}$.) $\mathcal{L}_0 \cap s(\mathcal{K})$ will be called the *indicative belief set,* and $\mathcal{L}_1 \cap s(\mathcal{K})$ the *$>$-extended belief set* (conditionals-extended belief set) that is associated with the belief state $\mathcal{K}$.

In most studies in the AGM tradition, belief states have been identified with sets of sentences that need not contain modal or conditional sentences. Belief states may then be identical to *indicative belief sets,* i.e., $\mathcal{K} = \mathcal{L}_0 \cap s(\mathcal{K})$. In discourse on the Ramsey test, it has—with the notable exception of Levi's writings—commonly been assumed that belief states are identical to *$>$-extended belief sets,* i.e., $\mathcal{K} = \mathcal{L}_1 \cap s(\mathcal{K})$. It has not always been sufficiently understood that $>$-extended belief sets differ in their logic, in essential respects, from indicative belief sets. Properties of the latter that are so self-evident that they are not even explicitly discussed may very well not hold of the former. Important examples of this are that a given indicative belief set typically has several proper subsets that are indicative belief sets, and that is also has several proper supersets that are indicative belief sets. In other words, for a typical belief state $\mathcal{K}$ there are (1) several consistent belief states $\mathcal{K}'$ such that $\mathcal{L}_0 \cap s(\mathcal{K}') \subset \mathcal{L}_0 \cap s(\mathcal{K})$, and (2) several consistent belief states $\mathcal{K}''$ such that $\mathcal{L}_0 \cap s(\mathcal{K}) \subset \mathcal{L}_0 \cap s(\mathcal{K}'')$. But the same does not hold for $>$-extended belief sets, i.e., it does not hold if $\mathcal{L}_0$ is replaced by $\mathcal{L}_1$ in the formulas.

To see why this is so, let us first consider the task of finding, for a given $>$-extended belief set, a proper subset that is also an $>$-

extended belief set. In other words, for a given belief state $\mathcal{K}$ we want to find a belief state $\mathcal{K}'$ such that $\mathcal{L}_1 \cap s(\mathcal{K}') \subset \mathcal{L}_1 \cap s(\mathcal{K})$. The corresponding task for indicative belief sets (i.e., to find some $\mathcal{K}'$ such that $\mathcal{L}_0 \cap s(\mathcal{K}') \subset \mathcal{L}_0 \cap s(\mathcal{K})$) can be solved by performing any nonvacuous contraction on $\mathcal{K}$.[15] (Unless, of course, $\mathcal{L}_0 \cap s(\mathcal{K}) = Cn(\varnothing)$.) This method does not in general work for >-extended belief sets, however, since a nonvacuous contraction typically provides support for conditional sentences that were not supported by the original belief state. If I contract my belief that John is mentally retarded, then I gain support for the conditional sentence 'If John has lived 30 years in London, then John understands the English language'. If I contract my belief that my wife has a headache, then I gain support for the sentence 'If it stops raining, then my wife would like to join me for an afternoon walk', etc.

Next let us consider the task of finding an >-extended belief set that is a proper superset of a given >-extended belief set. In other words, for a given belief state $\mathcal{K}$ we are searching for a belief state $\mathcal{K}''$ such that $\mathcal{L}_1 \cap s(\mathcal{K}) \subset \mathcal{L}_1 \cap s(\mathcal{K}'')$. The corresponding task for indicative belief sets (i.e., to find some $\mathcal{K}''$ such that $\mathcal{L}_0 \cap s(\mathcal{K}) \subset \mathcal{L}_0 \cap s(\mathcal{K}'')$) can be solved by expanding $\mathcal{K}$ by any indicative sentence p such that $\neg p \notin \mathcal{L}_0 \cap s(\mathcal{K})$. (Unless, of course, $s(\mathcal{K})$ is negation complete.) This method does not work for >-extended belief sets, however, since the addition of an indicative sentence that is compatible with all previously supported indicative sentences typically withdraws the support of conditional sentences that were previously supported. As an example of this, suppose that I previously knew nothing about John's profession. Furthermore, suppose that I acquire the belief that John is a driver by profession. No indicative sentence represents a belief that is lost by this,[16] but support will be lost for various conditional sentences, such as 'If John goes home from work by taxi every day, then he is a rich man'.

Thus, neither of the two obvious ways to obtain indicative belief sets K_1 and K_2 such that $K_1 \subset K_2$, namely, contraction and expansion, seems to work in general for >-extended belief sets. (As has been indicated by Levi and Rott, similarly formed $\Diamond$-extended belief sets, where $\Diamond$ denotes possibility, have essentially the same property. The nonvacuous contraction of q typically provides support for $\Diamond \neg q$, and the nonvacuous expansion by q withdraws support for $\Diamond \neg q$.[17])

---

[15] I use 'contraction' in the general sense of an operation on a belief state which retracts a certain sentence (or set of sentences) from the set of supported sentences. A contraction is nonvacuous if and only if the sentence to be retracted was supported by the original belief state.

[16] Except, of course, indicative sentences that refer to my belief state, such as 'I know nothing about John's profession'. The inclusion of such sentences into logically closed belief representations gives rise to serious problems. See my "The Paradox of the Believer," *Philosophia*, XXI (1991): 25–30. For our present purposes, these problems will be avoided by the assumption that neither $\mathcal{L}_0$ nor $\mathcal{L}_1$ contains doxastic or epistemic operators.

In order to see how the differences between indicative and >-extended belief sets will affect the proof of the impossibility theorem, let us consider the following version of the central part of Gärdenfors's proof. Following Gärdenfors, it will be assumed in the following passage that belief states are identical to the >-extended belief sets that they support.

> Let K be a >-extended belief set, let p and q be sentences such that $K + \neg q$, $K + (p \leftrightarrow q)$, and $K + (\neg p \,\&\, \neg q)$ are all >- extended belief sets and that $\neg p$ is an element only of the third of these. Then $K + \neg q$ and $K + (p \leftrightarrow q)$ are both proper subsets of $K + (\neg p \,\&\, \neg q)$. By (M), that follows from (R), we may then conclude that $(K + \neg q)^{*}p$ and $(K + (p \leftrightarrow q))^{*}p$ are both proper subsets of $(K + (\neg p \,\&\, \neg q))^{*}p$. Thus by (P) both $\neg q$ and $p \leftrightarrow q$ are elements of $K + (\neg p \,\&\, \neg q)^{*}p$. By (1), p is also an element of $K + (\neg p \,\&\, \neg q)^{*}p$, and by (2) $K + (\neg p \,\&\, \neg q)^{*}p$ is consistent. By (L), this is a contradiction.

By way of example, let us consider beliefs about a vote in parliament. Let p denote that a certain party is going to recommend its M.P.s to vote in favor of a certain bill, and let q denote that my M.P., Ms. Quinton, will vote in favor of that same bill. (For simplicity of logic, we shall disregard abstentions.) Let K be an >-extended belief set that corresponds to (and, on Gärdenfors's assumptions, is identical to) a belief state in which I know nothing about how the party or Ms. Quinton are going to vote.

In the proof (in the above version) it is assumed that, when I adopt (only) the belief 'Ms. Quinton is going to vote against the bill', then my new >-extended belief set will be $K + \neg q$. This, however, is not realistic. When learning that Ms. Quinton is going to vote against the bill, I also acquire support for the sentence 'If the party recommends voting for the bill, then Ms. Quinton votes contrary to the party's recommendation', i.e., $p > (q \leftrightarrow \neg p)$. This sentence is clearly not an element of $K + \neg q$.

Similarly, when I adopt (only) the belief 'Ms. Quinton is going to vote for the bill if and only if the party recommends its M.P.s to do so', then my >-extended belief set is assumed in the proof to be changed into $K + (p \leftrightarrow q)$. In this case, however, I also acquire support for the sentence 'If the party recommends voting for the bill, then Ms. Quinton votes in favor of the bill', i.e., $p > q$.

---

[17] Levi, "Iteration of Conditionals"; Rott, "Conditionals and Theory Change," p. 109.

$p > (q \blacksquare \neg p)$ and $p > q$ are clearly inconsistent. Therefore, the $>$-extended belief set that is obtained from K by acceptance of $\neg q$ and the $>$-extended belief set that is obtained from K by acceptance of $p \blacksquare q$ cannot both be subsets of one and the same consistent $>$-extended belief set (such as $K + (\neg p \,\&\, \neg q)$). Thus, the use of expansion in the proof is at variance with what seems to be a reasonable account of the dynamics of $>$-extended belief sets.

Admittedly, expansion and the (E) condition are not essential for the proof. In order to reconstruct the proof it is sufficient to construct, in some way or other, three $>$-extended belief sets K_1, K_2 and K such that $K_1 \subset K$, $K_2 \subset K$, $K_1 \not\subseteq K_2$ and $K_2 \not\subseteq K_1$. The most obvious way to do this would be to perform two different nonvacuous contractions of K. As was shown above, however, although contraction easily provides us with an indicative belief set that is a proper subset of a given indicative belief set, it does not in general provide us with a $>$-extended belief set that is a subset of a given $>$-extended belief set. Furthermore, a more general argument can be given which suggests that there is no way in which three $>$-extended belief sets K_1, K_2, and K with the desired properties can be obtained. The argument takes the form of a theorem that depends on the following four properties of belief states:

(LC) Let A be a set of sentences and p a sentence. If $A \subseteq s(\mathcal{H})$ and $p \in Cn(A)$, then $p \in s(\mathcal{H})$ (logical closure).

(NC) Let $\mathcal{H}$ be a consistent belief state and p a sentence such that $p \in s(\mathcal{H})$. Then $\neg p \notin s(\mathcal{H})$ (negation consistency).

(CC) Let $\mathcal{H}$ be a consistent belief state and p a consistent expression. If $p > q_1 \in s(\mathcal{H})$ and $p > q_2 \in s(\mathcal{H})$, then $q_1 \,\&\, q_2$ is consistent (conditional consistency).

(CP) For all p and q, if $p \in s(\mathcal{H})$, $q \notin s(\mathcal{H})$, and $\neg q \notin s(\mathcal{H})$, then $q > (p \,\&\, q) \in s(\mathcal{H})$ (conditional preservation).

Logical closure (LC) is a direct consequence of the meaning of 'support' in this context. It should be observed that closure is only postulated for the set of sentences that are supported by $\mathcal{H}$. The belief state $\mathcal{H}$ may nevertheless be a set of sentences that is not closed under consequence, or it need not be a set of sentences at all.

Negation consistency (NC) and conditional consistency (CC) may be seen less as postulates than as parts of a definition of what it means for belief states to be consistent. If the background logic is supraclassical, then (NC) and (CC) are necessary conditions for a belief state to be consistent. For my present purposes I do not need to know whether or not the conjunction of the two is a sufficient condition for consistency.

Conditional preservation (CP) is the least obvious of the four postulates. It can be argued for with more or less covert reference to the Ramsey test, but that would be too close to begging the question. It remains, therefore, to appeal directly to prephilosophical intuitions about conditional inference. Let p and q be arbitrary sentences that may be either connected or unconnected in terms of subject matter. It can easily be verified that (CP) conforms with intuition in various such cases. As one example among many, suppose that you believe that Stockholm is the capital of Sweden (p) but have no opinion of whether or not Stockholm is the largest city in Sweden (q). Then you believe that, if Stockholm is the largest city in Sweden, then it is both the capital of Sweden and that country's largest city. (As the reader can straightforwardly verify, the theorem also holds if $q > (p \,\&\, q)$ is replaced by $q > p$ in the definition of conditional preservation. $q > (p \,\&\, q)$ has been chosen since it is more conformant with linguistic convention.)

THEOREM 1. Let $\mathcal{K}$, $\mathcal{K}_1$, and $\mathcal{K}_2$ be consistent belief states that satisfy (LC), (NC), (CC), and (CP). Then, if $\mathcal{L}_1 \cap s(\mathcal{K}_1) \subseteq \mathcal{L}_1 \cap s(\mathcal{K})$ and $\mathcal{L}_1 \cap s(\mathcal{K}_2) \subseteq \mathcal{L}_1 \cap s(\mathcal{K})$, then either $\mathcal{L}_1 \cap s(\mathcal{K}_1) \subseteq \mathcal{L}_1 \cap s(\mathcal{K}_2)$ or $\mathcal{L}_1 \cap s(\mathcal{K}_2) \subseteq \mathcal{L}_1 \cap s(\mathcal{K}_1)$.

For proofs of theorems and observations the reader is referred to the appendix below.

In summary, the impossibility theorem depends on the unwarranted generalization of closure under expansion (E) from indicative belief sets, for which it holds, to >-extended belief sets, for which it does not hold. One way to reconstruct the theorem would be to use contraction in lieu of expansion to construct the three >-extended belief sets that are essential for the proof. That too, however, would require an unsustainable application to >-extended belief sets of principles that are valid for indicative belief sets. Indeed, if four reasonable postulates for the dynamics of belief states, namely, (LC), (NC), (CC), and (CP), are accepted, then it follows by theorem 1 that there is no way to reconstruct the proof of the impossibility theorem.

Having seen that the Ramsey test can be defended against the impossibility theorem, I shall now turn to the construction of a formal account of epistemic conditionals which accords with the Ramsey test.

III. A FORMAL ACCOUNT OF EPISTEMIC CONDITIONALS

In this and the following section, I shall make the simplifying assumption that the language contains no other symbol than > in addition to those of truth-functional and predicate logic, i.e., that $\mathcal{L}_1$ comprises the whole of the language.

According to the standard version of the Ramsey test, the set of conditional sentences that are supported by a belief state $\mathcal{H}$ can be determined from (1) the set of indicative sentences that $\mathcal{H}$ supports, and (2) the behavior of $\mathcal{H}$ under revision. Since belief states are then completely determined by these two factors, we may, without loss of generality, let each belief state be represented by a pair $\langle K, D\rangle$ of one indicative and one dynamic component. K, the indicative component, is a set of sentences such that $Cn(K) = \mathcal{L}_0 \cap s(\mathcal{H})$. K may, or may not, be closed under logical consequence so that $Cn(K) = K$. D, the dynamic component, is a formal structure of some kind that determines, for each sentence p, the outcome of the revision of $\langle K, D\rangle$ by p.

Due to our restriction of the language, $s(\langle K, D\rangle) \subseteq \mathcal{L}_1$ for all belief states $\langle K, D\rangle$. The support function s should have the following three properties:

(i) If $p \in \mathcal{L}_0$, then $p \in s(\langle K, D\rangle)$ iff $p \in Cn(K)$.
(ii) $s(\langle K, D\rangle) = Cn(s(\langle K, D\rangle))$
(iii) $p > q \in s(\langle K, D\rangle)$ holds if and only if $p > q$ is supported by the Ramsey test, as applied to $\langle K, D\rangle$.

According to (i), truth conditions (support conditions) for indicative sentences are independent of the dynamic component of the belief state. According to (ii), the support function is closed under logical consequence. (ii) implies, for instance, that if $p > q \in s(\langle K, D\rangle)$ and $r > t \in s(\langle K, D\rangle)$, then $((p > q) \,\&\, (r > t)) \in s(\langle K, D\rangle)$. (iii) signals compliance with the Ramsey test. Before more can be said about the formal interpretation of (iii), we need to have a closer look at the structure of the dynamic component D of a belief state $\langle K, D\rangle$.

The most obvious construction is to let D be either a revision operator for K or some formal device (such as a selection function or a relation of epistemic entrenchment) that uniquely determines a revision operator for K. In an important sense, however, revision operators such as those of the AGM formalism are insufficient for this purpose. When revising a belief state $\langle K, D\rangle$ by a sentence p, we want the result to be a new belief state with the same structure, $\langle K', D'\rangle$. Given a belief state $\langle K, *\rangle$, in which K is an indicative belief set and $*$ an AGM operator of revision for K, we obtain for each indicative sentence p a new indicative belief set $K' = K{*}p$. But we do not obtain, directly from K and $*$, a new revision operator $*'$ for K'. Since the dynamic component of the new belief state is not determined, neither is the set of conditional sentences that it supports.

There seem to be two major ways out of this dilemma. One of these is to find some formal structure for the dynamic component of $\langle K, D\rangle$ which does not have this limitation. In other words, for each indicative sentence p, D should determine not only the indicative belief set K^*p but also a new dynamic component D' that combines with K^*p to form a new belief state of the same form, $\langle K^*p, D'\rangle$. It is no easy matter, however, to find a plausible formal structure for the dynamic component which has this property.[18]

The other, more simplistic, solution is to assume that the dynamic component is determined by the indicative component, i.e., that for each K there is only one D with which it can be combined. This seems to have been commonly taken as a tacit assumption in the belief-dynamics literature. In an earlier paper,[19] I brought it up to the surface, giving it the following mathematical formulation: the background apparatus that is common to all belief states contains, along with the language and the logic, a function f, the *superselector* (so called since it selects selectors). To each set K that can form the indicative part of a belief state, f assigns a dynamic component $f(K)$. The latter may be either a revision operator for K or some selection mechanism (such as a selection function) that in its turn determines a revision operator for K. In the former case, given a belief state $\langle K, *\rangle$, revision by p results in the new belief state $\langle K^*p, f(K^*p)\rangle$

This construction clearly solves our dilemma from a formal point of view, and it provides us with an account of belief revision that treats first-order conditionals in the way prescribed by the Ramsey test. But the construction has a problem from the point of view of its intuitive interpretation: if it is assumed that the indicative components of belief states are closed under logical consequence, then for each indicative belief set there is, with this construction, only one revision operator with which it can be combined.

To see why this is an implausible feature of our construction, let p (as above) denote that the party recommends voting in favor of a certain bill and q that Ms. Quinton, M.P., votes in favor of it. Furthermore, let $K = Cn(\{p\})$, and let $*$ be a revision operator for K. Then, with weak and reasonable assumptions on $*$, $K^*q = K^*(p \equiv q)$, from which follows that $f(K^*q) = f(K^*(p \equiv q))$. It follows from this that the result of revising $\langle K, *\rangle$ first by q and then by $\neg p$ must be the same as that of revising $\langle K, *\rangle$ first by $p \equiv q$ and then by $\neg p$. From an intuitive point of view, however, the result of the first chain of operations should be a belief state that supports q, and that of the second chain of operations should be a belief state that supports $\neg q$.

---

[18] See my "In Defense of Base Contraction," in *Synthese*, XCI (1992): 239–45.

[19] "A Dyadic Representation of Belief," in Gärdenfors, ed., *Belief Revision* (New York: Cambridge, 1992), pp. 89–121

One would wish a formal system of belief dynamics to allow one and the same indicative belief set to take on different dynamic properties, and thus also to be combinable with different sets of conditional sentences. There is, perhaps surprisingly, a way to achieve this without giving up the formal simplicity of the superselector approach outlined above. The trick is to let K, the indicative part of the belief state, be a belief base that is not closed under logical consequence. One and the same indicative belief set can then be represented by different belief bases, each of which has the belief set as its logical closure. The superselector will assign different revision operators to different belief bases with the same indicative belief set. In this way, one and the same indicative belief set can appear in belief states that are different in terms of their dynamic properties and in terms of the conditional sentences that they support.

To see how this works in the above example, let the indicative belief set $Cn(\{p\})$ of the original belief state be represented by the belief base $\{p\}$. With reasonable assumptions about (minimal) revisions of belief bases, the result of revising the original belief state by q should be a belief state with the belief base $\{p, q\}$, and further revision by $\neg p$ should result in a belief state with the belief base $\{\neg p, q\}$. On the other hand, revision by $p \equiv q$ should yield a belief state with the belief base $\{p, p \equiv q\}$, and further revision by $\neg p$ should yield the belief base $\{\neg p, p \equiv q\}$. Thus, q is supported in the first case and $\neg q$ in the second, in conformance with what was called for in the informal example.

The congruity in this case between intuitive inference and operations on belief bases is no coincidence. As I have indicated elsewhere,[20] the belief base approach can account for a surprising range of cases in which one and the same indicative belief set appears in belief states with different dynamic properties. (The usefulness of the belief base approach for the analysis of conditional sentences was pointed out by M. Morreau.[21])

In summary, the application of superselectors to belief sets allows for a formal account of belief dynamics that

(1) conforms with the Ramsey test, and
(2) provides, as the result of a revision, not only a new indicative belief set but also a new $>$-extended belief set.

---

[20] "New Operators For Theory Change," *Theoria*, LV (1989). 114–132; and "In Defense of Base Contraction."

[21] "Epistemic Semantics for Counterfactuals," *Journal of Philosophical Logic*, XXI (1992): 33–62.

If we use belief bases instead of belief sets, then the formal system also

(3) allows one and the same indicative belief set to be the indicative part of different $>$-extended belief sets.

I do not wish to claim that there is no other way to achieve these properties, but this is a simple and plausible way that is well worth exploring.

Since, in this approach, the second (dynamic) component of a belief state is a function of the first (indicative) component, namely, the belief base, we may omit reference to the second component. In other words, each belief state can be identified with its belief base. This will be done in the rest of this paper.

We now have enough background for the formal statement of a belief-revision system that conforms with the Ramsey test for (first-order) conditionals. Revision operators will be based on selection functions in the way that is standard in the AGM tradition.[22]

DEFINITION 1. Let $K \subseteq \mathcal{L}_0$ and $q \in \mathcal{L}_0$. $K \perp q$, the *remainder* of K by q, is the set of inclusion-maximal subsets of K that do not have q as a logical consequence.

A *selection function* for K is a function γ such that for all $q \in \mathcal{L}_0$: $\gamma(K \perp q)$ is a nonempty subset of $K \perp q$, unless the latter is empty, in which case $\gamma(K \perp q) = \{K\}$.

A selection function γ for K is *relational* if and only if there is a relation $\preccurlyeq$ on the set of subsets of K such that for all nonempty $K \perp p$:

$$\gamma(K \perp p) = \{K' \in K \perp p \mid K'' \preccurlyeq K' \text{ for all } K'' \in K \perp p\}.$$

DEFINITION 2. $*$ is an operator of *partial meet revision* for a subset K of $\mathcal{L}_0$ if and only if there is a selection function γ for K such that for all $p \in \mathcal{L}_0$: $K{*}p = (\cap\gamma(K \perp \neg p)) \cup \{p\}$

DEFINITION 3. A *belief-revision system* (in a given language and relative to a given logic) is a pair $\langle \mathcal{B}, f \rangle$, where $\mathcal{B}$ is a set of belief bases and f is a function that to each $K \in \mathcal{B}$ assigns a revision operator $*_K$ such that for all indicative sentences p of the language, $K{*_K}p \in \mathcal{B}$.[23]

A belief-revision system is *orderly* if and only if for each $K \in \mathcal{B}$, $f(K)$ is a partial meet revision operator that is based on a selection function that is relational by some relation $\preccurlyeq_K$ that is reflexive, transitive, and complete, and also has the maximizing property, namely, that for its strict counterpart $\prec$, if $A \subset B$ then $A \prec B$.

---

[22] Cf. My "Reversing the Levi Identity," forthcoming in *Journal of Philosophical Logic.*

[23] I owe the basic idea for this to a suggestion by André Fuhrmann in correspondence.

> If $\langle \mathcal{B}, f \rangle$ is orderly, then the (global) revision operator $*$ such that for each $K \in \mathcal{B}$, $K^*p = K^{*}{}_{K}p$ (with $*_K = f(K)$) is an *orderly revision operator*.

The support condition (truth condition) for indicative sentences has already been mentioned: if $p \in \mathcal{L}_0$, then $p \in s(K)$ if and only if $p \in Cn(K)$. For conditional sentences, the following support condition will be used:

> (R1) Let p and q be indicative sentences and K a set of indicative sentences. Then $p > q \in s(K)$ iff $q \in K^*p$, where $*$ is the unique revision operator that is assigned to K in the belief-revision model.

This account of conditional sentences is "parasitic" in Levi's terminology, i.e., the set of supported conditional sentences is (in a given belief-revision system) completely determined by a set of supported indicative sentences (the belief base).[24] The above argument may be seen as a demonstration that a parasitic account of conditionals can be motivated independently of Levi's metaphysical standpoint on which sentences can or cannot be true or false.

IV. ITERATED CONDITIONALS

An important limitation of this type of parasitic account was pointed out by Gärdenfors when commenting on Levi's analysis of conditionals: "The main problem with this position is that it does not account for *iterated* conditionals. What kind of belief would a proposition like the following be on this view? 'If this vase breaks if dropped on the floor, then it breaks if thrown against the wall'."[25] In an answer to Gärdenfors, Levi questioned that "bona fide examples of such iterated conditionals are used in serious discourse or inquiry,"[26] but he nevertheless offered acceptability criteria for some types of iterated conditionals. He did not, however, provide a fully formalized criterion for conditionals of the type $(p > q) > (r > t)$, and he "cheerfully plead[s] guilty" to the charge that "this account of iterated conditionals is not fully systematic in the sense that I have failed to supply a single acceptability condition applicable to both noniterated and iterated conditionals" (*ibid.*, p. 78). According to Levi, both $p > q$ and $r > t$ in this example are disposition sentences of a type that should not be interpreted as conditionals.

---

[24] Cf. Levi, "Iteration of Conditionals," p. 66.

[25] Gärdenfors, "Belief Revisions," p. 90

[26] Levi, "Iteration of Conditionals," p. 73.

In Levi's approach, conditional sentences are parasitic on (logically closed) indicative belief sets (corpora), rather than on belief bases. Gärdenfors's argument, however, is equally valid against the approach that was introduced in section III, with its parasitism on belief bases. In this section, I intend to show that there is a reasonable interpretation of the Ramsey test under which a fully systematic account can be obtained of iterated conditionals of arbitrary complexity.

The following is the most general version of the new interpretation:

> (R2) For any belief state $\mathcal{K}$ and any two sentences p, $q \in \mathcal{L}_1$, $p > q \in s(\mathcal{K})$ iff $q \in s(K')$ for any $\mathcal{K}'$ that is the outcome of a smallest possible change of $\mathcal{K}$ to make it support p.

The major difference between (R1) and (R2) is that, whereas the former stipulates a specified method by which to incorporate p into the belief state, the latter only postulates that the outcome of the change (however it is brought about) should be that p is supported and that, given this, the original belief state has been changed as little as possible. One argument in favor of the more general approach of (R2) is that, in some cases, the most natural way to "add" a sentence to one's stock of supported sentences does not seem to be an operation of revision. In particular, as was observed by Levi, the process of acquiring support for the sentence $\Diamond p$ ('it is possible that p') seems to coincide with the process of *contracting* the sentence $\neg p$ (*ibid.*, pp. 70–1).

In what follows, (R2) will be applied to belief states of the type that we ended up with in the previous section; thus, belief states are represented by belief bases. In order to give precision to the notion of a smallest possible change (or minimal change), I shall make use of three-termed similarity relations of the type proposed by David Lewis:[27]

$S(X, Y, Z)$ "X is at least as similar to Y as is Z to Y"

The strict counterpart will here be denoted as follows:

$\hat{S}(X, Y, Z) = S(X, Y, Z) \mathbin{\&} \neg S(Z, Y, X)$

Lewis proposed that the similarity relation should be a (weak) ordering and that each world should resemble itself more than any other world resembles it (*ibid.*, pp. 48–9). Thus, S should satisfy the following conditions:

---

[27] *Counterfactuals* (Cambridge: Blackwell, 1973)

(S1) $S(X, W, Y) \vee S(Y, W, X)$ completeness
(S2) $S(X, W, X)$ reflexivity
(S3) $S(X, W, Y) \,\&\, S(Y, W, Z) \rightarrow S(X, W, Z)$ transitivity
(S4) $S(X, W, W) \rightarrow X = W$ centering

For each world W, axioms S1–S4 completely characterize those sphere systems around W in which all worlds belong to some sphere (*ibid.*, p. 49). These axioms can also be derived from plausible conditions on a more general format for similarity, namely, a four-termed similarity relation 'X is at least as similar to Y as Z is similar to W'.[28]

In his seminal work on similarity of sets, A. Tversky[29] proposed that the similarity between two sets A and B should be measured by a real-valued function $F(A \cap B, A\backslash B, B\backslash A)$ that increases with its first argument and decreases with its second and last arguments. The following two conditions on a similarity relation follow his ideas so closely that I venture to call them the *Tversky properties:*

(T1) If $K' \cap K \subset K'' \cap K$ and $K''\backslash K = K'\backslash K$, then $\hat{S}(K'', K, K')$.
(T2) If $K' \cap K = K'' \cap K$ and $K''\backslash K \subset K'\backslash$K, then $\hat{S}(K'', K, K')$.

The Tversky properties make (S4) superfluous:

OBSERVATION 1. If a similarity relation S satisfies (S2), (S3), (T1), and (T2), then it satisfies (S4).

We now have the tools necessary to give precision to the phrase 'smallest possible change' of (R2) in terms of similarity relations:

DEFINITION 4. K is a p-set iff either $p \in K$ or $p \in s(K)/\mathcal{L}_0$.
K' is a (*consistent*) *K-closest p-set* by a similarity relation S iff K' is a consistent p-set and there is no consistent p-set K'' such that $\hat{S}(K'', K, K')$.

For simplicity, the word 'consistent' will be deleted in the phrase 'consistent K-closest p-set'.

---

[28] Cf T. Williamson, "First-Order Logics for Comparative Similarity," *Notre Dame Journal of Formal Logic*, XXIX (1988): 457–81; my "Similarity Semantics and Minimal Changes of Belief," forthcoming in *Erkenntnis*.

[29] "Features of Similarity," *Psychological Review*, LXXXIV (1977): 327–52.

(R2′) For any belief base K and any two sentences p, $q \in \mathcal{L}_1$, $p > q \in s(K)$ iff $q \in s(K')$ for all K' that are K-closest p-sets by a given similarity relation S.

This will be called the ***similarity-based Ramsey test,*** in contrast to the (conventional) version (R1), which will be called a ***revision-based Ramsey test.*** The revision-based version has the important advantage of being simple and direct for first-order conditionals. The similarity-based version has the advantage of being defined for iterated conditionals of arbitrary complexity. In order to combine both advantages, one would need a similarity-based Ramsey test that coincides for first-order conditionals with a revision-based Ramsey test. In other words, the following type of relationship should hold between a similarity relation and a revision operator:

DEFINITION 5. A similarity relation S *corresponds* to an operator $*$ if and only if for each consistent set K of indicative sentences and each consistent indicative sentence p, $K{*}p$ is the intersection of the set of K-closest p-sets by p.

It follows from the definition that for each similarity relation S there is exactly one operator $*$ such that S corresponds to $*$. The following theorem shows that we can indeed combine the advantages of the two versions of the Ramsey test. In other words, the parasitic account can be extended to cover iterated conditionals of arbitrary complexity.

THEOREM 2. Let $*$ be an orderly revision operator. Then there is a similarity relation S that corresponds to $*$ and that satisfies (S1), (S2), (S3), (S4), (T1), and (T2).

In general, there are many ways in which an orderly revision operator can be "extended" to a similarity relation to which it corresponds. In other words, revision by conditionals requires more information about entrenchment or epistemic value than what is required for revision by indicatives.

The following are some of the logical properties of the conditional connective in this framework:

OBSERVATION 2. Let K be a consistent subset of $\mathcal{L}_0$, and let $>$ be evaluated, in the manner of (R2′), according to a similarity relation that satisfies (S1), (S2), (S3), (S4), (T1), and (T2). Then

(1) $p > p \in s(K)$
(2) $p > q \in s(K)$ iff $p > p \;\&\; q \in s(K)$

(3) If $p > q_k \in s(K)$ for all $1 \leq k \leq n$ and $r \in Cn(\{q_1, \dots q_n\})$, then $p > r \in s(K)$.

The following relationships *do not hold* in general, however:

(n1) If $(p \vee q) > r \in s(K)$ then $p > r \in s(K)$ (SDA, simplification of disjunctive antecedents).
(n2) If $p > r \in s(K)$ then $p \;\&\; q > r \in s(K)$ (strengthening the antecedent).
(n3) If $p > q \in s(K)$ then $\neg q > \neg p \in s(K)$.
(n4) If $p > q \in s(K)$ and $q > r \in s(K)$, then $p > r \in s(K)$ (transitivity).

SVEN OVE HANSSON
Uppsala University

APPENDIX: PROOFS

Before the proofs of the theorems and observations, two lemmas will be given that are needed for the proof.

LEMMA 1. If a similarity relation S satisfies (S3), then it satisfies the following properties·

(1) $\hat{S}(X, W, Y) \;\&\; S(Y, W, Z) \rightarrow \hat{S}(X, W, Z)$
(2) $S(X, W, Y) \;\&\; \hat{S}(Y, W, Z) \rightarrow \hat{S}(X, W, Z)$

PROOF OF LEMMA 1. For part 1, suppose that $\hat{S}(X, W, Y)$ and $S(Y, W, Z)$. It follows by (S3) that $S(X, W, Z)$. If $S(Z, W, X)$, then it follows by (S3) from $S(Y, W, Z)$ and $S(Z, W, X)$ that $S(Y, W, X)$, contrary to $\hat{S}(X, W, Y)$. Thus $\neg S(Z, W, X)$. It follows from $S(X, W, Z)$ and $\neg S(Z, W, X)$ that $\hat{S}(X, W, Z)$.

For part 2, suppose that $S(X, W, Y)$ and $\hat{S}(Y, W, Z)$. It follows by (S3) that $S(X, W, Z)$. If $S(Z, W, X)$, then it follows by (S3) from $S(Z, W, X)$ and $S(X, W, Y)$ that $S(Z, W, Y)$, contrary to $\hat{S}(Y, W, Z)$. Thus $\neg S(Z, W, X)$. It follows from $S(X, W, Z)$ and $\neg S(Z, W, X)$ that $\hat{S}(X, W, Z)$.

LEMMA 2. Let $p \in \mathcal{L}_0$. If K' is a K-closest p-set by a similarity relation S that satisfies quasi transivity ($\hat{S}(X, W, Y) \;\&\; \hat{S}(Y, W, Z) \rightarrow \hat{S}(X, W, Z)$), (T1) and (T2), then $K' = M \cup \{p\}$ for some $M \in K \perp \neg p$.

PROOF OF LEMMA 2. Let K' be a K-closest p-set. First suppose that $K' \cap K \notin K \perp \neg p$. Then, since $\neg p \notin Cn(K')$, there is some $M \in K \perp \neg p$ such that $K' \cap K \subset M \in K \perp \neg p$. Let $K'' = M \cup \{p\}$. Then $K' \cap K \subset K'' \cap K$. It follows by (T1) that $\hat{S}(K'', K, (K' \cap K) \cup \{p\})$. If $K' \backslash K \subseteq \{p\}$, then $\hat{S}(K'', K, K')$ follows directly. If $K' \backslash K \not\subseteq \{p\}$, then (T2) yields $\hat{S}((K' \cap K) \cup \{p\}, K, K')$, and quasi transitivity yields $\hat{S}(K'', K, K')$. It follows from $\hat{S}(K'', K, K')$ that K' cannot be a K-closest p-set, contrary to the conditions. We may conclude from this contradiction that $K' \cap K \in K \perp \neg p$.

Next, suppose that $K' \backslash K \not\subseteq \{p\}$. Let $K'' = (K' \cap K) \cup \{p\}$. We then have $K' \cap K = K'' \cap K$ and $K'' \backslash K \subset K' \backslash K$. It follows by (T2) that $\hat{S}(K'', K, K')$, so that K' cannot be a K-closest p-set, contrary to the conditions. Thus $K' \backslash K \subseteq \{p\}$, from which the lemma follows directly.

PROOF OF THEOREM 1. Suppose not. Then there are x_1 and x_2 in $\mathcal{L}_1$ such that $x_1 \in s(\mathcal{K}_1) \backslash s(\mathcal{K}_2)$ and $x_2 \in s(\mathcal{K}_2) \backslash s(\mathcal{K}_1)$.

We are first going to show that $(\neg x_1 \vee \neg x_2) \notin s(\mathcal{K}_1)$. Suppose not. Then it follows by $x_1 \in s(\mathcal{K}_1)$ and the logical closure (LC) of $s(\mathcal{K}_1)$ that $\neg x_2 \in s(\mathcal{K}_1)$. Thus $\neg x_2 \in s(\mathcal{K})$, contrary to $x_2 \in s(\mathcal{K}_2) \subseteq s(\mathcal{K})$ and the negation consistency (NC) of $s(\mathcal{K})$. We may conclude that $(\neg x_1 \vee \neg x_2) \notin s(\mathcal{K}_1)$. Furthermore, from $x_2 \notin s(\mathcal{K}_1)$ follows, by the logical closure (LC) of $s(\mathcal{K}_1)$, that $\neg(\neg x_1 \vee \neg x_2) \notin s(\mathcal{K}_1)$. By conditional preservation (CP) follows that $((\neg x_1 \vee \neg x_2) > (x_1 \,\&\, (\neg x_1 \vee \neg x_2))) \in s(\mathcal{K}_1)$, i.e., that $((\neg x_1 \vee \neg x_2) > (x_1 \,\&\, \neg x_2)) \in s(\mathcal{K}_1)$.

By a symmetrical argument follows $((\neg x_1 \vee \neg x_2) > (x_2 \,\&\, \neg x_1)) \in s(\mathcal{K}_2)$. By $s(\mathcal{K}_1) \subseteq s(\mathcal{K})$ and $s(\mathcal{K}_2) \subseteq s(\mathcal{K})$ follows that both

$$((\neg x_1 \vee \neg x_2) > (x_1 \,\&\, \neg x_2)) \text{ and } ((\neg x_1 \vee \neg x_2) > (x_2 \,\&\, \neg x_1))$$

are elements of $\mathcal{K}$.

Suppose that $\neg x_1 \vee \neg x_2$ is inconsistent. Then x_2 is a logical truth so that, by logical closure (LC), $x_2 \in s(\mathcal{K}_1)$. Since this is not the case, we may conclude that $\neg x_1 \vee \neg x_2$ is consistent. Thus conditional consistency (CC) is violated. This contradiction completes the proof.

PROOF OF THE THEOREM 2. For each set K, let $\preceq_K$ be the maximizing ordering on the subsets of K on which revisions of K by $*$ are based. Let u be any real-valued function u such that $u(\emptyset) = 0$, $u(K) < 1$ and for all M, $N \subseteq K$: $u(M) \leq u(N)$ iff $M \preceq_K N$. For all sets A, let $|A|$ denote the number of elements of A. The following construction will be used:

$$S(M, K, N) \text{ iff } |M \backslash K| - u(M \cap K) \leq |N \backslash K| - u(N \cap K)$$

It follows straightforwardly that (S1), (S2), (S3), (S4), (T1), and (T2) are satisfied. We are going to show that S corresponds to $*$. For any indicative sentence p, let $Z(K, p) = \{M \cup \{p\} \mid M \in K \perp \neg p$ and $S(M \cup \{p\}, K, N \cup \{p\})$ for all $N \in K \perp \neg p.\}$. We are going to show (I) that $Z(K, p)$ is the set of K-closest p-sets by S, and (II) that $K{*}p = \bigcap Z(K, p)$.

Part I. If K' is a K-closest p-set by S, then it follows from lemmas 1 and 2 that $K' \in Z(K, p)$. It remains to be shown that if $K' \in Z(K, p)$, then K' is a K-closest p-set by S. Suppose to the contrary that $K' \in Z(K, p)$ and that K' is not a K-closest p-set by S. Then there is some consistent set K'' such that $p \in K''$ and $\hat{S}(K'', K, K')$, i.e., $|K'' \backslash K| - u(K'' \cap K) < |K' \backslash K| - u(K' \cap K)$. Let K_1 be any set such that $K'' \cap K \subseteq K_1 \in K \perp \neg p$, and let $K_2 = K_1 \cup \{p\}$. Then $K_2 \backslash K \subseteq K'' \backslash K$, so that $|K_2 \backslash K| \leq |K'' \backslash K|$. By the reflexivity and maximizing property of $\preceq_K$ it follows from

$K'' \cap K \subseteq K_1 = K_2 \cap K$ that $K'' \cap K \preceq_K K_2 \cap K$, so that $u(K'' \cap K) \leq u(K_2 \cap K)$. It follows that

$$|K_2 \backslash K| - u(K_2 \cap K) \leq |K'' \backslash K| - u(K'' \cap K),$$

so that $|K_2 \backslash K| - u(K_2 \cap K) < |K' \backslash K| = u(K' \cap K)$, i.e., $\hat{S}(K_2, K, K')$, contrary to $K' \in Z(K, p)$. It follows from this contradiction that $Z(K, p)$ is the set of K-closest p-sets by S.

Part II. By the definitions,

$$K{*}p = (\cap\gamma(K \perp \neg p) \cup \{p\})$$

$$= (\cap\{M \in K \perp \neg p \mid N \preceq_K M \text{ for all } N \in K \perp \neg p\} \cup \{p\}).$$

From $M, N \in K \perp \neg p$ it follows that $p \in M$ iff $p \in N$, so that $|(M \cup \{p\}) \backslash K| = |(N \cup \{p\}) \backslash K|$. Therefore, $S((M \cup \{p\}), K, (N \cup \{p\}))$ holds iff $u((N \cup \{p\}) \cap K) \leq u((M \cup \{p\}) \cap K)$. Since p is an element of both M and N iff it is an element of K, this is equivalent to $u(N) \leq u(M)$, and thus to $N \preceq_K M$. We may therefore substitute $S((M \cup \{p\}), K, (N \cup \{p\}))$ for $N \preceq_K M$ in the above formula for $K{*}p$, and obtain: $K{*}p = (\cap\{M \in K \perp \neg p \mid S(M \cup \{p\}, K, N \cup \{p\})$ for all $N \in K \perp \neg p\} \cup \{p\}) = \cap Z(K, p)$. This concludes the proof of the theorem.

PROOF OF OBSERVATION 1. Suppose that (S2), (S3), (T1), and (T2) are satisfied and that $X \neq W$. We are going to show that $\neg S(X, W, W)$.

From $X \neq W$ it follows that either $W \backslash X \neq \varnothing$ or $X \backslash W \neq \varnothing$.

If $W \backslash X \neq \varnothing$, then $\hat{S}(W, W, X \cap W)$ follows by (T1). If $X \subseteq W$ then $S(X \cap W, W, X)$ follows by (S2), and otherwise it follows by (T2). It then follows by (S3) and lemma 1 that $\hat{S}(W, W, X)$.

If $X \backslash W \neq \varnothing$, then $\hat{S}(X \cap W, W, X)$ follows by (T2). If $W \subseteq X$, then $S(W, W, X \cap W)$ follows by (S2), and otherwise it follows by (T1). It then follows by (S3) and lemma 1 that $\hat{S}(W, W, X)$.

In both cases, $\hat{S}(W, W, X)$ yields $\neg S(X, W, W)$.

PROOF OF OBSERVATION 2. The proofs of (1)–(3) are simple.

For (n1), let $K = \{q, r, p \rightarrow \neg r\}$, and let S be such that $\{p, q, p \rightarrow \neg r\}$ is the K-closest p-set by S.

For (n2), let $K = \{p, r, q \rightarrow \neg r\}$, and let S be such that $\{p, p \,\&\, q, q \rightarrow \neg r\}$ is the K-closest $p \,\&\, q$-set by S.

For (n3), let $K = \{p, q\}$, and let S be such that $\{p, \neg q\}$ is the K-closest $\neg q$-set by S.

For (n4), let $K = \{\neg p, q, r, p \rightarrow \neg r\}$, and let S be such that $\{p, q, p \rightarrow \neg r\}$ is the K-closest p-set by S.

NEW BOOKS

ADLER, JACOB: *The Urgings of Conscience: A Theory of Punishment.* Philadelphia: Temple, 1992. ix, 316 p. Cloth $44.95.

AGGER, BEN: *The Discourse of Domination: From the Frankfurt School to Postmodernism.* Evanston: Northwestern, 1992 xii, 347 p. Cloth $42.95, paper $17.95.

ALMEDER, ROBERT: *Blind Realism: An Essay on Human Knowledge and Natural Science.* Savage, MD: Rowman & Littlefield, 1992. xiii, 247 p. Cloth $47.50.

ANDERSON, ALAN ROSS, NUEL D. BELNAP, JR., and J. MICHAEL DUNN: *Entailment: The Logic of Relevance and Necessity, Vol. II.* Princeton: University Press, 1992. xxvii, 749 p. Cloth $75.00.

ANNAS, JULIA E.: *Hellenistic Philosophy of Mind.* Berkeley: California UP, 1992. ix, 245 p. Cloth $35.00.

APPIAH, KWAME ANTHONY: *In My Father's House: Africa in the Philosophy of Culture.* New York: Oxford, 1992. xi, 225 p. Cloth $29.95.

BARRETT, CYRIL: *Wittgenstein on Ethics and Religious Belief.* Cambridge: Blackwell, 1991. xiv, 285 p. Cloth $44.95.

BEAMISH, ROB: *Marx, Method, and the Division of Labor.* Champaign: Illinois UP, 1992. x, 196 p. Cloth $24.95.

BEINER, RONALD: *What's the Matter with Liberalism?* Berkeley: California UP, 1992. viii, 197 p. Cloth $25.00.

BEISER, FREDERICK C.: *Enlightenment, Revolution, and Romanticism: The Genesis of Modern German Political Thought, 1790–1800.* Cambridge: Harvard, 1992. xiii, 434 p. Cloth $45.00.

BELL, ALBERT A., JR. and JAMES B. ALLIS: *Resources in Ancient Philosophy: An Annotated Bibliography of Scholarship in English, 1965–1989.* Metuchen, NJ: Scarecrow, 1991. xvii, 799 p. Cloth $79.50.

BELLIOTTI, RAYMOND A.: *Justifying Law: The Debate over Foundations, Goals, and Methods.* Philadelphia: Temple, 1992. ix, 301 p. Cloth $49.95.

BERNSTEIN, J. M.: *The Fate of Art.* University Park: Penn State UP, 1992. x, 292 p. Cloth $45.00, paper $14.95.

BERNSTEIN, MARK H.: *Fatalism.* Lincoln: Nebraska UP, 1992. xi, 165 p. Cloth $25.00.

BERNSTEIN, RICHARD J.: *The New Constellation: The Ethical-Political Horizons of Modernity/Post-Modernity.* Cambridge: MIT, 1992. 358 p. Cloth $35.00, paper $15.95.

BHASKAR, ROY: *Philosophy and the Idea of Freedom.* Cambridge: Blackwell, 1991. xii, 202 p. Cloth $47.95, paper $19.95.

BILGRAMI, AKEEL: *Belief and Meaning.* Cambridge: Blackwell, 1992. ix, 301 p. Cloth $54.95.

BLANCHETTE, OLIVIA: *The Perfection of the Universe according to Aquinas: A Teleological Cosmology.* University Park: Penn State UP, 1992. xvii, 334 p. Cloth $35.00.

BORGMANN, ALBERT: *Crossing the Postmodern Divide.* Chicago: University Press, 1992. vii, 173 p. Cloth $19.95.

BOURDIEU, PIERRE, and LOIC J. D. WACQUANT: *An Invitation to Reflexive Sociology*. Chicago: University Press, 1992. xiv, 332 p. Cloth $38 95, paper $13.95.

BOVE, PAUL A.: *Mastering Discourse: The Politics of Intellectual Discourse*. Durham: Duke, 1992. xiv, 252 p. Cloth $45.00, paper $16.95.

BRADLEY, RAYMOND: *The Nature of All Being: A Study of Wittgenstein's Modal Atomism*. New York: Oxford, 1992. xxi, 244 p. Cloth $39.95.

BRANDT, RICHARD B.: *Morality, Utilitarianism, and Rights*. New York: Cambridge, 1992. viii, 393 p. Cloth $49.95, paper $17.95.

BRODY, HOWARD: *The Healer's Power*. New Haven: Yale, 1991. xiii, 311 p. Cloth $30.00.

BROMBERGER, SYLVAIN: *On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*. Chicago: University Press and Stanford: CSLI, 1992. vii, 231 p. Cloth $41.95, paper $16.95.

BROOME, JOHN: *Weighing Goods: Equality, Uncertainty, and Time*. Cambridge: Blackwell, 1991. x, 255 p. Cloth $54.95.

BRUMBAUGH, ROBERT S.: *Western Philosophic Systems and Their Cyclic Transformations*. Carbondale: Southern Illinois UP, 1992. xvi, 156 p. Cloth $29.95.

CAMPBELL, JAMES: *The Community Reconstructs: The Meaning of Pragmatic Social Thought*. Urbana: Illinois UP, 1992. ix, 147 p. Cloth $27.50, paper $12.95.

CARRUTHERS, PETER: *Human Knowledge and Human Nature: A New Introduction to an Ancient Debate*. New York: Oxford, 1992. viii, 199 p. Cloth $55.00.

CASCARDI, ANTHONY J.: *The Subject of Modernity*. New York: Cambridge, 1992. x, 316 p. Cloth $49.95, paper $16.95.

CHARLTON, WILLIAM: *The Analytic Ambition: An Introduction to Philosophy*. Cambridge: Blackwell, 1991. x, 217 p. Cloth $47.95, paper $19.95.

CHURCHLAND, PATRICIA S., and TERRENCE J. SEJNOWSKI: *The Computational Brain*. Cambridge: MIT, 1992. xi, 544 p. Cloth $39.95.

COADY, C. A. J.: *Testimony: A Philosophical Study*. New York: Oxford, 1992 x, 315 p. Cloth $39.95.

COHEN, JEAN L., and ANDREW ARATO: *Civil Society and Political Theory*. Cambridge: MIT, 1992. xxi, 771 p. Cloth $45.00.

COLEMAN, JANET: *Ancient and Medieval Memories: Studies in the Reconstruction of the Past*. New York: Cambridge, 1992. xx, 646 p. Cloth $85.00.

COLES, ROMAND: *Self/Power/Other: Political Theory and Dialogical Ethics*. Ithaca: Cornell, 1992. x, 203 p. Cloth $29.95.

CONNOLLY, WILLIAM E.: *Identity/Difference: Democratic Negotiations of Political Paradox*. Ithaca: Cornell, 1992. xii, 244 p. Cloth $27.50.

CRAVEN, JOHN: *Social Choice: A Framework for Collective Decisions and Individual Judgements*. New York: Cambridge, 1992. x, 152 p. Cloth $49.95, paper $15.95.

DAWSON, DOYNE: *Cities of the Gods: Communist Utopias in Greek Thought*. New York: Oxford, 1992. viii, 305 p. Cloth $45.00.

DE LANDA, MANUEL: *War in the Age of Intelligent Machines*. New York: Zone, 1991. viii, 272 p. Cloth $32.95, paper $16.95.

DETLEFSEN, MICHAEL: *Proof, Logic, and Formalization.* New York: Routledge, 1992. x, 241 p. Cloth $52.50.

DILMAN, ILHAM: *Philosophy and the Philosophic Life: A Study in Plato's* Phaedo. New York: St. Martin's, 1991. xii, 139 p. Cloth $45.00.

DONNER, WENDY: *The Liberal Self: John Stuart Mill's Moral and Political Philosophy.* Ithaca: Cornell, 1992. x, 225 p. Cloth $35.00, paper $12.95.

EAGLETON, TERRY, FREDERIC JAMESON, and EDWARD W. SAID: *Nationalism, Colonialism, and Literature.* Minneapolis: Minnesota UP, 1991. 102 p. Cloth $25.00, paper $9.95.

EARMAN, JOHN: *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge: MIT, 1992. xiv, 272 p. Cloth $35.00.

EASTON, PATRICIA, THOMAS M. LENNON, and GREGOR SEBBA: *Bibliographia Malebranchia: A Critical Guide to the Malebranche Literature into 1989.* Carbondale: Southern Illinois UP, 1992. xxiii, 189 p. Cloth $24.95.

ECO, UMBERTO, with RICHARD RORTY, JONATHAN CULLER, and CHRISTINE BROOKE-ROSE: *Interpretation and Overinterpretation.* STEPHAN COLLINI, ed. New York: Cambridge, 1992. ix, 151 p. Cloth $39.95, paper $11.95.

EELLS, ELLERY: *Probabilistic Causality.* New York: Cambridge, 1991. xii, 413 p. Cloth $44.50.

FELDMAN, FRED: *Confrontations with the Reaper: A Philosophical Study of the Nature and Value of Death.* New York: Oxford, 1992. xiv, 249 p. Cloth $24.95.

FERRARA, LAWRENCE: *Philosophy and the Analysis of Music: Bridges to Musical Sound, Form, and Reference.* Westport, CT: Greenwood, 1991. xxii, 360 p. Cloth $45.00.

FERREIRA, KEITH N.: *The Intellectual Rebel.* New York: Vantage, 1992. v, 73 p. Cloth $7.95.

FIUMARA, GEMMA CORRADI: *The Symbolic Function: Psychoanalysis and the Philosophy of Language.* Cambridge: Blackwell, 1992. vii, 230 p. Cloth $44.95, paper $21.95.

FLYNN, BERNARD: *Political Philosophy at the Closure of Metaphysics.* Atlantic Highlands, NJ: Humanities, 1992. vii, 234 p. Cloth $45.00.

FOGELIN, ROBERT J.: *Philosophical Interpretations.* New York: Oxford, 1992. 255 p. Cloth $29.95.

FORESTER, TOM, and PERRY MORRISON: *Computer Ethics: Cautionary Tales and Ethical Dilemmas in Computing.* Cambridge: MIT, 1992. viii, 193 p. Paper $12.50.

FRIEDMAN, MICHAEL: *Kant and the Exact Sciences.* Cambridge: Harvard, 1992. xvii, 357 p. Cloth $45.00.

FROHLICH, NORMAN, and JOE A. OPPENHEIMER: *Choosing Justice: An Experimental Approach to Ethical Theory.* Berkely: California UP, 1992. xiv, 258 p. Cloth $45.00.

GANS, CHAIM: *Philosophical Anarchism and Political Disobedience.* New York: Cambridge, 1992. xiii, 174 p Cloth $44.95.

GARBER, DANIEL: *Descartes' Metaphysical Physics.* Chicago: University Press, 1992. xiv, 389 p. Cloth $60.00, paper $23.95.

GAVIN, WILLIAM JOSEPH: *William James and the Reinstatement of the Vague.* Philadelphia: Temple, 1992. xi, 227 p. Cloth $37.95.

GELLNER, ERNEST: *Reason and Culture*. Cambridge: Blackwell, 1992. xi, 193 p. Cloth $44.95, paper $17.95.

GILBERT, MARGARET: *On Social Facts*. Princeton: University Press, 1992. x, 521 p. Cloth $69.50, paper $19.95.

GOEHR, LYDIA: *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. New York: Oxford, 1992. vii, 314 p. Cloth $59.00.

GOLDBLATT, ROBERT: *Logics of Time and Computation:* 2nd Ed., revised and expanded. Stanford: CSLI, 1992. ix, 180 p. Cloth $39.95, paper $16.95.

GRAFSTEIN, ROBERT: *Institutional Realism: Social and Political Constraints on Rational Actors*. New Haven: Yale, 1992. xii, 244 p. Cloth $30.00.

GRAHAM, KEITH: *Karl Marx, Our Contemporary: Social Theory for a Post-Leninist World*. Toronto: University Press, 1992. x, 182 p. Cloth $50.00, paper $17.95.

GREEN, RONALD M.: *Kierkegaard and Kant: The Hidden Debt*. Albany: SUNY, 1992. xviii, 301 p. Cloth $49.50, paper $16.95.

GREENAWALT, KENT: *Law and Objectivity*. New York: Oxford, 1992. x, 288 p. Cloth $45.00.

GROVER, DOROTHY: *A Prosentential Theory of Truth*. Princeton: University Press, 1992. xii, 288 p. Cloth $45.00.

GRUNDMANN, REINER: *Marxism and Ecology*. New York: Oxford, 1992. x, 324 p. Cloth $67.00.

GUEST, STEPHEN: *Ronald Dworkin*. Stanford: University Press, 1991. ix, 320 p. Cloth $42.50, paper $14.95.

HAMMOND, MICHAEL, JANE HOWARTH, and RUSSELL KEAT: *Understanding Phenomenology*. Cambridge: Blackwell, 1991. viii, 319 p. Cloth $52.95, paper $20.95.

HANNOSH, MICHELE: *Baudelaire and Caricature: From the Comic to an Art of Modernity*. University Park: Penn State UP, 1992. xii, 348 p. Cloth $39.50.

HARRE, ROM: *Physical Being: A Theory for Corporeal Psychology*. Cambridge: Blackwell, 1991. ix, 265 p. Cloth $49.95.

HARRIS, ERROL E.: *Spinoza's Philosophy: An Outline*. Atlantic Highlands, NJ: Humanities, 1992. x, 125 p. Cloth $35.00, paper $12.50.

HAUSMAN, DANIEL M.: *The Inexact and Separate Science of Economics*. New York: Cambridge, 1992. xi, 372 p. Cloth $59.95.

HAYRY, HETA: *The Limits of Medical Paternalism*. New York: Routledge, 1992. ix, 208 p. Cloth $59.50, paper $23.95.

HETHERINGTON, STEPHEN CADE: *Epistemology's Paradox: Is a Theory of Knowledge Possible?* Savage, MD: Rowman & Littlefield, 1992. x, 234 p. Cloth $46.75.

HEYD, DAVID: *Genethics: Moral Issues in the Creation of People*. Berkeley: California UP, 1992. xiii, 276 p. Cloth $45.00.

HILL, THOMAS E., JR.: *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell, 1992. xi, 255 p. Cloth $36.50, paper $13.95.

HIRSCHMANN, NANCY J.: *Rethinking Obligation: A Feminist Method for Political Theory*. Ithaca: Cornell, 1992. xi, 365 p. Cloth $39.95, paper $15.95.

HOLUB, RENATE: *Antonio Gramsci: Beyond Marxism and Postmodernism*. New York: Routledge, 1992. xii, 247 p. Cloth $55.00, paper $16.95.

HOLUB, ROBERT C: *Crossing Borders: Reception Theory, Poststructuralism, Deconstruction.* Madison: Wisconsin UP, 1992. xii, 244 p. Cloth $42.00, paper $16.95.

INNES, JULIE: *Privacy, Intimacy, and Isolation.* New York: Oxford, 1992. ix, 157 p. Cloth $24.95.

JEFFREY, RICHARD: *Probability and the Art of Judgement.* New York: Cambridge, 1992. xi, 244 p. Cloth $49.95, paper $16.95.

KAHAN, ALAN S.: *Aristocratic Liberalism: The Social and Political Thought of Jacob Burckhardt, John Stuart Mill, and Alexis de Tocqueville.* New York: Oxford, 1992. viii, 228 p. Cloth $39.95.

KALUPAHANA, DAVID J.: *A History of Buddhist Philosophy: Continuities and Discontinuities.* Honolulu: Hawaii UP, 1992. xvi, 304 p. Cloth $36.00, paper $14.95.

KAMM, FRANCES MYRNA: *Creation and Abortion: A Study in Moral and Legal Philosophy.* New York: Oxford, 1992. viii, 227. Cloth $29.95, paper $11.95.

KEANEY, JOHN J.: *The Composition of Aristotle's* Athenaion Politeia: *Observation and Explanation.* New York: Oxford, 1992. xii, 191 p. Cloth $39.95.

KELLY, GEORGE ARMSTRONG: *The Humane Comedy: Constant, Toqueville, and French Liberalism.* New York: Cambridge, 1992. xiv, 262 p. Cloth $59.95.

KILNER, JOHN F.: *Who Lives? Who Dies? Ethical Criteria in Patient Selection.* New Haven: Yale, 1992. xiv, 359 p. Cloth $35.00, paper $15.00.

KING, RICHARD H.: *Civil Rights and the Idea of Freedom.* New York: Oxford, 1992. xii, 269 p. Cloth $35.00.

KLEINBERG, STANLEY S.: *Politics and Philosophy: The Necessity and Limitations of Rational Argument.* Cambridge: Blackwell, 1991. ix, 213 p. Cloth $44.95, paper $19.95.

KLOSKO, GEORGE: *The Principle of Fairness and Political Obligation.* Savage, MD: Rowman & Littlefield, 1992. xiii, 204 p. Cloth $52.50, paper $19.95.

KOONS, ROBERT: *Paradoxes of Belief and Strategic Rationality.* New York: Cambridge, 1992. xi, 174 p. Cloth $49.95.

KOSLOW, ARNOLD: *A Structuralist Theory of Logic.* New York: Cambridge, 1992. xi, 418 p Cloth $69.50.

KVANVIG, JONATHAN L.: *The Intellectual Virtues and the Life of the Mind.* Savage, MD: Rowman & Littlefield, 1991. xii, 198 p. Cloth $39.50.

LEVINSON, HENRY SAMUEL: *Santayana, Pragmatism, and the Spiritual Life.* Chapel Hill: North Carolina UP, 1992. xvi, 348 p. Cloth $39.95.

LEWIS, FRANK A.: *Substance and Predication in Aristotle.* New York: Cambridge, 1992. xvii, 370 p. Cloth $44.50.

LEYTON, MICHAEL: *Symmetry, Causality, Mind.* Cambridge: MIT, 1992. vii, 630 p. Cloth $45.00.

LLOYD, S. A.: *Ideals as Interests in Hobbes'* Leviathan: *The Power of Mind over Matter.* New York: Cambridge, 1992. xi, 396 p. Cloth $54.95.

LOUDEN, ROBERT B.: *Morality and Moral Theory.* New York: Oxford, 1992. xiii, 230 p. Cloth $35.00, paper $15.95.

LUKE, TIMOTHY: *Shows of Force: Power, Politics, and Ideology in Art Exhibitions*. Durham: Duke, 1992. x, 250 p. Cloth $37.95, paper $15.95.

MCKENNA, ANDREW J.: *Violence and Difference: Girard, Derrida, and Deconstruction*. Champaign: Illinois UP, 1992. xi, 238 p. Cloth $39.95, paper $15.95.

MARSH, JAMES, JOHN D. CAPUTO, and MEROLD WESTPHAL: *Modernity and Its Discontents*. New York: Fordham, 1992. xiv, 219 p. Cloth $32.50, paper $19.95.

MARTIN, BILL: *Matrix and Line: Derrida and the Possibilities of Postmodern Social Theory*. Albany, SUNY, 1992. xii, 255 p. Cloth $49.50, paper $16.95.

MARTIN, JULIAN: *Francis Bacon, the State, and the Reform of Natural Philosophy*. New York: Cambridge, 1992. xiii, 236 p. Cloth $49.95

MARTINICH, A. P.: *The Two Gods of* Leviathan: *Thomas Hobbes on Religion and Politics*. New York: Cambridge, 1992. xiv, 430 p. Cloth $59.95.

MATTEO, ANTHONY M.: *Quest for the Absolute: The Philosophical Vision of Joseph Marechal*. Dekalb: Northern Illinois UP, 1992. xi, 173 p. Cloth $30.00.

MELE, ALFRED: *Springs of Action: Understanding Intentional Behavior*. New York: Oxford, 1992 ix, 272 p. Cloth $39.95.

MERRELL, FLOYD: *Sign, Textuality, World*. Bloomington: Indiana UP, 1992. xviii, 264 p. Cloth $37.50.

MIDGLEY, MARY: *Science as Salvation: A Modern Myth and its Meaning*. New York: Routledge, 1992. x, 239 p. Cloth $25.00.

MILLER, BARRY: *From Existence to God: A Contemporary Philosophical Argument*. New York: Routledge, 1992. x, 206 p. Cloth $42.50.

MILLER, RICHARD W.: *Moral Differences: Truth, Justice, and Conscience in a World of Conflict*. Princeton: University Press, 1992. ix, 396 p. Cloth $65.00, paper $16.95.

MORRIS, BRIAN: *Western Conceptions of the Individual*. New York· Berg, 1992. ix, 505 p. Cloth $74.00.

NATHAN, N. M. L.: *Will and World: A Study in Metaphysics*. New York: Oxford, 1992. x, 171 p. Cloth $49.95.

NEVILLE, ROBERT CUMMINGS: *God the Creator: On the Transcendence and the Presence of God*. Albany: SUNY, 1992. xxv, 320 p. Paper $12.95.

NICHOLS, MARY P.: *Citizens and Statesmen: A Study of Aristotle's* Politics. Savage, MD: Rowman & Littlefield, 1991. ix, 233 p. Cloth $50.00, paper $19.95.

NICHOLSON, GRAEME: *Illustrations of Being: Drawing upon Heidegger and upon Metaphysics*. Atlantic Highlands, NJ: Humanities, 1992. xiii, 293 p. Cloth $49.95.

OAKLEY, JUSTIN: *Morality and the Emotions*. New York: Routledge, 1992. x, 253 p. Cloth $49.95.

O'HARA, DANIEL T.: *Radical Parody: American Culture and Critical Agency after Foucault*. New York: Columbia, 1992. xvi, 311 p. Cloth $37.50.

OLSON, ALAN M.: *Hegel and the Spirit: Philosophy as Pneumatology*. Princeton: University Press, 1992. xiv, 223 p. Cloth $24.95.

O'NEILL, JOHN: *Worlds without Content: Against Formalism.* New York: Routledge, 1992. x, 189 p. Cloth $49.95.

PASSMORE, JOHN: *Serious Art.* La Salle, IL: Open Court, 1991. x, 301 p. Cloth $49.95, paper $19.95.

PELIKAN, JAROSLAV: *The Idea of the University: A Re-examination.* New Haven. Yale, 1992. x, 238 p. Cloth $30.00.

PENELHUM, TERRENCE: *David Hume: An Introduction to His Philosophical System.* West Lafayette: Purdue, 1992. xv, 218 p. Cloth $27.00, paper $13.75.

PERRY, MICHAEL J.: *Love and Power: The Role of Religion and Morality in American Politics.* New York: Oxford, 1991. viii, 218 p. Cloth $29.95.

PIKE, NELSON: *Mystic Union: An Essay in the Phenomenology of Mysticism.* Ithaca: Cornell, 1992. xiv, 224 p. Cloth $29.95.

PLOCHMANN, GEORGE KIMBALL: *Richard McKeon: A Study.* Chicago: University Press, 1990. xvi, 260 p. Cloth $29.95.

POLS, EDWARD: *Radical Realism: Direct Knowing in Science and Philosophy.* Ithaca: Cornell, 1992. xi, 221 p. Cloth $28.50.

PURDY, LAURA M.. *In Their Best Interest? The Case against Equal Rights for Children.* Ithaca: Cornell, 1992. vii, 262 p. Cloth $32.50, paper $12.95.

PUTNAM, HILARY: *Realism with a Human Face.* JAMES CONANT, ed. Cambridge: Harvard, 1992. lxxiv, 347 p. Cloth $37.50, paper $15.95.

RASMUSSEN, DOUGLAS B., and DOUGLAS J. DEN UYL: *Liberty and Nature: An Aristotelian Defense of Liberal Order.* La Salle, IL: Open Court, 1991. xvi, 268 p. Cloth $49.95, paper $24.95.

REIMAN, JEFFREY: *Justice and Modern Moral Philosophy.* New Haven: Yale, 1992. xiv, 322 p. Cloth $35.00, paper $16.00.

ROBERTS, JULIAN: *The Logic of Reflection: German Philosophy in the Twentieth Century.* New Haven: Yale, 1992. viii, 307 p. Cloth $30.00.

ROBINSON, WILLIAM S.: *Computers, Minds, and Robots.* Philadelphia: Temple, 1992. x, 281 p. Cloth $39.95.

ROCKMORE, TOM: *Irrationalism: Lukacs and the Marxist View of Reason.* Philadelphia: Temple, 1992. 315 p. Cloth $44.95. And *On Heidegger's Philosophy and Nazism.* Berkeley: California UP, 1992. xi, 393 p. Cloth $45.00.

ROSEMONT, HENRY, JR.: *A Chinese Mirror: Moral Reflections on Political Economy and Society.* La Salle, IL: Open Court, 1991. x, 130 p. Cloth $27.95, paper $9.95.

ROTENSTREICH, NATHAN: *Immediacy and Its Limits: A Study in Martin Buber's Thought.* Philadelphia: Harwood, 1992. 118 p. Cloth $20.00.

RYAN, MARIE-LAURE: *Possible Worlds, Artificial Intelligence, and Narrative Theory.* Bloomington: Indiana UP, 1992. viii, 291 p. Cloth $35.00.

SCHAUER, FREDERICK: *Playing by the Rules: A Philosophical Examination of Rule-based Decision-making in Law and in Life.* New York: Oxford, 1992. xvii, 254 p. Cloth $39.95.

SCHEFFLER, SAMUEL: *Human Morality.* New York: Oxford, 1992. 150 p. Cloth $26.00.

SCHMITT, FREDERICK F.: *Knowledge and Belief.* New York: Routledge, 1992. x, 278 p. Cloth $49.95.

SCHOFIELD, MALCOLM: *The Stoic Idea of the City.* New York: Cambridge, 1992. xii, 164 p. Cloth $44.95.

SCHROEDER, FREDERIC M.: *Form and Transformation: A Study in the Philosophy of Politics.* Montreal: McGill-Queen's, 1992. xiv, 125 p. Cloth $34.95.

SEAGER, WILLIAM: *Metaphysics of Consciousness.* New York: Routledge, 1992. viii, 262 p. Cloth $45.00.

SEARLE, JOHN R.: *The Rediscovery of the Mind.* Cambridge: MIT, 1992. xv, 270 p. Cloth $22.50.

SEYHAN, AZADE: *Representation and Its Discontents: The Critical Legacy of German Romanticism.* Berkeley: California UP, 1992. ix, 187 p. Cloth $35.00, paper $16.00.

SHERRY, PATRICK: *Spirit and Beauty: An Introduction to Theological Aesthetics.* New York: Oxford, 1992. viii, 192 p. Cloth $49.95.

SHERWIN, SUSAN: *No Longer Patient: Feminist Ethics and Health Care.* Philadelphia: Temple, 1992. xi, 286 p. Cloth $39.95.

SHRADER-FRECHETTE, K. S.: *Risk and Rationality: Philosophical Foundations for Populist Reforms.* Berkeley: California UP, 1992. x, 312 p. Cloth $39.95, paper $15.95.

SIMMONS, A. JOHN: *The Lockean Theory of Rights.* Princeton: University Press, 1992. ix, 387 p. Cloth $39.50.

SMILEY, MARION: *Moral Responsibilities and the Boundaries of Community: Power and Accountability from a Pragmatic Point of View.* Chicago: University Press, 1992. x, 286 p. Cloth $44.50, paper $17.95.

SMITH, A. J.: *Metaphysical Wit.* New York: Cambridge, 1992. xii, 270 p. Cloth $54.95.

SMITH, JOHN E.: *America's Philosophical Vision.* Chicago: University Press, 1992. vii, 213 p. Cloth $39.95, paper $15.95.

SNARE, FRANCIS. *The Nature of Moral Thinking.* New York: Routledge, 1992. vi, 187 p. Cloth $49.95, paper $14.95.

SOKOLOWSKI, ROBERT: *Pictures, Quotations, and Distinctions: Fourteen Essays in Phenomenology.* Notre Dame: University Press, 1992. xiv, 325 p. Cloth $32.95.

SPECTOR, HORACIO: *Autonomy and Rights: The Moral Foundations of Liberalism.* New York: Oxford, 1992. 196 p. Cloth $49.95.

STAUMBAUGH, JOAN: *The Finitude of Being.* Albany: SUNY, 1992. viii, 200 p. Cloth $44.50, paper $14.95.

STRAWSON, P. F.: *Analysis and Metaphysics: An Introduction to Philosophy.* New York: Oxford, 1992. viii, 144 p. Cloth $42.50, paper $12.95.

STRUEVER, NANCY S.: *Theory as Practice: Ethical Inquiry in the Renaissance.* Chicago: University Press, 1992. xiii, 246 p. Cloth $32.50.

SUPPES, PATRICK: *Language for Humans and Robots.* Cambridge: Blackwell, 1992. xiv, 417 p. Cloth $44.95.

SWANSON, JUDITH A.: *The Public and the Private in Aristotle's Political Philosophy.* Ithaca: Cornell, 1992. xiv, 244 p. Cloth $32.95.

SWANTON, CHRISTINE: *Freedom: A Coherence Theory*. Indianapolis: Hackett, 1992. x, 200 p. Cloth $29 50.

SWINBURNE, RICHARD: *Revelation: From Metaphor to Analogy*. New York: Oxford, 1992. 236 p. Cloth $65.00, paper $19.95.

TEICHMANN, ROGER: *Abstract Entities*. New York: St Martin's, 1992. xi, 177 p. Cloth $49.95.

THAGARD, PAUL: *Conceptual Revolutions*. Princeton: University Press, 1992. xvi, 285 p. Cloth $35.00.

TUANA, NANCY: *Woman and the History of Philosophy*. New York: Paragon, 1992. xv, 158 p. Paper $14.95.

WALBRIDGE, JOHN: *The Science of Mystic Lights: Qutb al-Din Shirazi and the Illuminationist Tradition in Islamic Philosophy*. Cambridge: Harvard, 1992. xvii, 296 p. Paper $14.95.

WEISBERG, RICHARD: *Poethics and Other Strategies of Law and Literature*. New York: Columbia, 1992. xvi, 312 p. Cloth $32.95.

WEISS, PAUL: *Creative Ventures*. Carbondale: Southern Illinois UP, 1992. xiv, 341 p. Cloth $45.00.

WENZ, PETER S.: *Abortion Rights as Religious Freedom*. Philadelphia: Temple, 1992. xiii, 294 p. Cloth $49.95, paper $22.95.

WESTON, ANTHONY: *Toward Better Problems: New Perspectives on Abortion, Animal Rights, the Environment, and Justice*. Philadelphia: Temple, 1992. xiii, 214 p. Cloth $44.95, paper $18.95.

WETZEL, JAMES: *Augustine and the Limits of Virtue*. New York: Cambridge, 1992. xv, 246 p. Cloth $54.95.

WHITE, STEPHEN K.: *Political Theory and Post-modernism*. New York: Cambridge, 1991. xiv, 153 p. Cloth $39.50, paper $12.95.

WILLIAMS, MICHAEL: *Unnatural Doubts*. Cambridge: Blackwell, 1992. xxiii, 386 p. Cloth $44.95.

WOLGAST, ELIZABETH: *Ethics of an Artificial Person: Lost Responsibility in Professions and Organizations*. Stanford: University Press, 1992. x, 161 p. Cloth $29.50, paper $10.95.

WOLIN, RICHARD: *The Terms of Cultural Criticism: The Frankfurt School, Existentialism, Poststructuralism*. New York: Columbia, 1992. xxi, 256 p. Cloth $35.00.

ZAMMITO, JOHN H.: *The Genesis of Kant's* Critique of Judgment. Chicago: University Press, 1992. ix, 479 p. Cloth $65.00, paper $18.95

ZEMACH, EDDY M.: *The Reality of Meaning and the Meaning of Reality*. Hanover, NH: New England UP, 1992. vii, 220 p. Cloth $45.00.

NEW BOOKS: ANTHOLOGIES

ANDERSON, GORDON L., and MORTON A. KAPLAN, eds.: *Morality and Religion*. New York: Paragon, 1992. viii, 363 p. Cloth $34.95.

BENSON, HUGH H., ed.: *Essays on the Philosophy of Socrates*. New York: Oxford, 1992. xv, 361 p. Cloth $45.00, paper $15.95.

BROWNING, DON S., and FRANCIS SCHUSSLER FIORENZA, eds.: *Habermas, Modernity, and Public Theology*. New York: Crossroad, 1992. vi, 258 p. Cloth $24.95.

BURKHOLDER, LESLIE, ed.: *Philosophy and the Computer*. Boulder: Westview, 1992. xi, 268 p. Cloth $55.00.

BUTLER, JUDITH, and JOAN W. SCOTT, eds.: *Feminists Theorize the Political*. New York: Routledge, 1992. xvii, 485 p. Cloth $55.00, paper $16 95.

CARMAN, JOHN, and MARK JUERGENSMEYER, eds.: *A Bibliographic Guide to the Comparative Study of Ethics*. New York: Cambridge, 1991. vii, 811 p Cloth $100.00.

DREYFUS, HUBERT L., and HARRISON HALL, eds.: *Heidegger: A Critical Reader*. Cambridge: Blackwell, 1992. vii, 303 p. Cloth $44.95.

GÄRDENFORS, PETER, ed.: *Belief Revision*. New York: Cambridge, 1992. 277 p. Cloth $34.95.

GRIERE, RONALD N., ed.: *Minnesota Studies in the Philosophy of Science, Vol. XV: Cognitive Models of Science*. Minneapolis: Minnesota UP, 1992. xxviii, 508 p. Cloth $44.95.

HARVEY, ELIZABETH D., and KATHLEEN OKRUHLIK, eds.: *Women and Reason*. Ann Arbor: Michigan UP, 1992. x, 304 p. Cloth $39.50.

HONNETH, AXEL, et al., eds.: *Philosophical Interventions in the Unfinished Project of Enlightenment*. Cambridge: MIT, 1992. x, 351 p Cloth $37.50, paper $17.95.

KIVY, PETER, ed.: *Essays on the History of Aesthetics*. Rochester: University Press, 1992. xi, 410 p. Cloth $65.00.

LASLETT, PETER, and JAMES S. FISHKIN, eds.: *Justice between Age Groups and Generations*. New Haven: Yale, 1992. viii, 243 p. Cloth $30.00.

LAWSON, BILL E., ed : *The Underclass Question*. Philadelphia: Temple, 1992. xv, 214 p. Cloth $34.95.

LEVINE, GEORGE, ed.: *Constructions of the Self*. New Brunswick· Rutgers, 1992. viii, 300 p. Cloth $45.00, paper $18.00.

MCCARTHY, GEORGE, ed.: *Marx and Aristotle: Nineteenth-Century German Social Theory and Classical Antiquity*. Savage, MD: Rowman & Littlefield, 1992. x, 379 p. Cloth $55.00, paper $22.00.

MCMULLIN, ERNAN, ed.: *The Social Dimensions of Science*. Notre Dame: University Press, 1992. xvii, 299 p. Cloth $37.95, paper $19.95.

PAUL, ELLEN FRANKEL, FRED D. MILLER, JR., and JEFFREY PAUL, eds.: *Economic Rights*. New York: Cambridge, 1992. xiv, 314 p. Paper $19.95.

PETERSON, GRETE B., ed.: *The Tanner Lectures on Human Values, Vol. XIII*. Salt Lake City: Utah UP, 1992. ix, 358 p. Cloth $30.00.

RORTY, AMÉLIE OKSENBERG, ed.: *Essays on Aristotle's* Poetics. Princeton: University Press, 1992. xii, 435 p. Cloth $69.50, paper $19.95.

SCHARLEMANN, ROBERT P., ed.: *Negation and Theology*. Charlottesville: Virginia UP, 1992. 155 p. Cloth $30.00.

WARNER, MARTIN, ed.: *Religion and Philosophy*. New York: Cambridge, 1992. xi, 155 p. Paper $17.95.

WARTENBERG, THOMAS E., ed.: *Rethinking Power*. Albany: SUNY, 1992. xxiii, 351 p. Cloth $49.50, paper $16.95.

WOOD, DAVID, ed.: *Derrida: A Critical Reader*. Cambridge: Blackwell, 1992. ix, 297 p. Cloth $54.95, paper $19.95; and *On Paul Ricoeur: Narrative and Interpretation*. New York: Routledge, 1992. xiii, 209 p. Cloth $67.50, paper $17.95.

THE JOURNAL OF PHILOSOPHY

is pleased to announce the publication of its

TWENTY-FIVE YEAR INDEX,
1964–1988

VOLUME LXI TO VOLUME LXXXV

AUTHORS, SUBJECTS, & REVIEWS

Supplement to its

FIFTY-YEAR INDEX, 1904–1953

AND

TEN-YEAR SUPPLEMENT, 1954–1963

$25.00	Cloth—Libraries
20.00	Cloth—Individuals
10.00	Paper—Individuals
8.00	Paper—Individuals (with new subscriptions)

Please send a check or money order to:

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through May 31, 1993. Orders and payment must be received by then. Paper copies will be mailed immediately, but cloth copies will take 6–8 weeks for delivery.

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709-M Philosophy Hall, Columbia University
New York, New York 10027

The Journal of Philosophy and Hackett Publishing Co.
invite you to review these distinguished recipients of the

JOHNSONIAN PRIZE IN PHILOSOPHY

ROBERT G. MUEHLMANN

Berkeley's Ontology

1992 448 pp. $37.95 cloth prof. price $25.50
awarded the 1991 prize

This provocative new work takes a more sharply focused look at Berkeley's ontology than has hitherto been attempted. Along the way, Muehlmann challenges some well-established interpretations of Berkeley's philosophy.

CHRISTINE SWANTON

Freedom: A Coherence Theory

210 pp. $29.50 cloth prof. price $18 00
awarded the 1990 prize

"The best philosophical work on freedom that I have ever read. Swanton does for the concept of freedom what Rawls did for the concept of justice."

—R. Hursthouse, *The Open University*

MORELAND PERKINS

Sensing the World

362 pp. $16.50 paper prof. price $12.50
awarded the 1982 prize

"A marvelous book... It is the most powerful, philosophically sophisticated argument for indirect realism I have ever read. It deserves to be at the center of contemporary discussion."

—Fred Dretske, *Stanford University*

HARRY van der LINDEN

Kantian Ethics & Socialism

384 pp. $16.50 paper prof. price $12.50
awarded the 1985 prize

"Throws valuable light on the three greatest modern German political thinkers: Kant, Hegel, and Marx."

—Patrick Riley, *University of Wisconsin*

VANN McGEE

Truth, Vagueness, & Paradox: An Essay on the Logic of Truth

246 pp. $34.95 cloth prof. price $20.00
awarded the 1988 prize

"McGee's own account of truth is not only a significant technical achievement in logic, but also a philosophical achievement, offering a new way of looking at the putatively universal notion of truth (or truth in our language)."

—Charles Parsons, *Harvard University*

C. L. HARDIN

Color for Philosophers: Unweaving the Rainbow

Foreword by Arthur Danto
268 pp. $16.00 paper prof. price $12.50
awarded the 1986 prize

"Much the best philosophically oriented book about colour that has been written."

—Jonathan Westphal, *Mind*

ROBERT DUNN

The Possibility of Weakness of Will

185 pp. $14.50 paper prof. price $9.50
awarded the 1984 prize

"Reveals... a mastery of the literature, a lucidity of mind, a rigor of argument, and a robust sense of psychological realism. It soars yet rarely loses sight of the terrain."

—*Canadian Philosophical Review*

CHARLES GUIGNON

Heidegger & the Problem of Knowledge

269 pp. $12.75 paper prof. price $9.50
honorable mention 1981

Hackett Publishing Co.
Box 44937, Indianapolis, IN 46244-0937
Telephone 317 635-9250 Fax 317 635-9292

Oxford

Regulating Toxic Substances

A Philosophy of Science and the Law

CARL F. CRANOR,
University of California, Riverside

"An extraordinary contribution in uncovering the ethical and normative questions hidden in the regulatory debate about hazardous substances....This book will help both to bring clarity to the debate about toxic substances and restore ethical reasoning to its appropriate central position in public discourse about environmental hazards."—Donald A. Brown, Esq., *Director of the Department of Environmental Resources, Commonwealth of Pennsylvania*

(Environmental Ethics and Science Policy Series)

1992 272 pp.; 8 illus. $39.95

Working Without a Net

A Study of Egocentric Epistemology

RICHARD FOLEY,
Rutgers University

In this new book, Foley defends an epistemology that takes seriously the perspectives of individual thinkers. He argues that having rational opinions is a matter of meeting our own internal standards rather than standards that are somehow imposed upon us from the outside.

1992 256 pp. $35.00

The Mind Matters

Consciousness and Choice in a Quantum World

DAVID HODGSON, *Judge of the Supreme Court, New South Wales*

"A worthwhile contribution to an important debate. It is well written, his philosophy is sound, and the description of quantum physics one of the best ever offered for the non-specialist."—*The Independent*

1991 496 pp.; 22 illus. $89.00

Winner of the 1992 Bertrand Russell Society Book Award and Chosen as an Outstanding Academic Book for 1991-1992 by Choice

Russell's Idealist Apprenticeship

NICHOLAS GRIFFIN,
McMaster University

The first detailed study of the early period of Russell's philosophical career, this book examines Russell's philosophical education at Cambridge in the early 1890s and his conversion to neo-Hegelianism; and the problems that ultimately led him to reject neo-Hegelianism.

1991 424 pp. $98.00

New in paperback!

Natural Reasons

Personality and Polity

S.L. HURLEY,
St. Edmund's Hall, Oxford

"An important, densely argued, ingenious book about the problems of choice among conflicting values and about how the concept of a person influences and is influenced by rational decisions about those values."—*The Philosophical Quarterly*

1989 (paper 1992) 480 pp.; 31 illus.
paper $18.95 cloth $49.95

Mysticism and Language

Edited by STEVEN T. KATZ,
Cornell University

Each of the ten essays included in this volume analyzes the relationship between language and mystical experience in one of the world's great religious traditions. The essays are by some of the foremost authorities in the field, including Ninian Smart, William Alston, Moshe Idel, Bernard McGinn, Carl Ernst, Ewert Cousins, Steven Katz, Bernard Faure, Stephen Phillips, and Bimal Matilal.

1992 272 pp. $38.00

Aristotle on the Perfect Life

ANTHONY KENNY,
Oxford University

Kenny offers a fresh examination and interpretation of the *Eudemian Ethics* on the basis of the 1991 Oxford Classical Text. He combines scholarly discussion of the Greek texts with reflection of the topics covered by Aristotle, taking account of post-Aristotelian treatments of themes such as moral vocation and moral luck.

1992 184 pp. $49.95

In the Interest of the Governed

A Study in Bentham's Philosophy of Utility and Law

Revised Edition

DAVID LYONS, *Cornell University*

This book offers a reinterpretation of Bentham's main philosophical doctrines, his principle of utility and his analysis of law, philosophical doctrines, as they are developed in Bentham's most important works.

1991 176 pp. $39.95

New in paperback!

Irrationality

An Essay on Akrasia, Self-Deception, and Self-Control

ALFRED R. MELE,
Davidson College

"A clear and careful analysis.... Rigorously argumentative and will be appreciated not only by philosophers specially concerned with the problems of akratic action, but also by those with a general interest in intentionality and the philosophy of action."—*Review of Metaphysics*

1987 (paper 1992) 200 pp.
paper $14.95

Perspectives on Thomas Hobbes

Edited by G.A.J. ROGERS,
University of Keele, and
ALAN RYAN, *Princeton University*

Including contributions from such notable scholars as David Gauthier, Noel Malcolm, D.D. Raphael, and Richard Tuck, this volume offers a variety of views on central aspects of Hobbes's life and work.

(Mind Association Occasional Series)

1989 (paper 1991) 224 pp.
paper $19.95 cloth $45.00

Hare and Critics

Essays on Moral Thinking

Edited by DOUGLAS SEANOR,
Georgia State University, and
N. FOTION, *Emory University*
With comments by R.M. HARE

This collection of thirteen original essays discusses the philosophy of R.M. Hare put forth in his book *Moral Thinking*, including his thoughts on universalizability, moral psychology, and the role of common-sense moral principles.

1988 (paper 1991) 320 pp.
paper $19.95 cloth $64.00

New in paperback!

Identity, Consciousness and Value

PETER UNGER,
New York University

"Its wealth of detail, its rigorous argumentation, and its large-scale deployment of the contemporary tools of discussion will ensure it a respected place in the literature....A full, clear, and relentless book."—*Times Literary Supplement*

1990 (paper 1992) 368 pp.
paper $19.95 cloth $42.00

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MG
To order by phone using major credit cards please call 1-800-451-7556

Oxford University Press
200 Madison Avenue ◆ New York, NY 10016

Chicago

Notebooks for an Ethics
Jean-Paul Sartre

Presented here for the first time in English, *Notebooks for an Ethics* is Sartre's attempt to develop an ethics consistent with the profound individualism of his existential philosophy. He grapples anew here with such central issues as "authenticity" and the relation of alienation and freedom to moral values.
Cloth $49.95 608 pages

The Genesis of Kant's *Critique of Judgment*
John H. Zammito

Zammito reveals the major transformations that the Third Critique underwent before publication and suggests that the work was animated by Kant's personal rivalry with Herder and by his commitment to Christian ideas of God and human moral freedom.

"A work of extraordinary erudition." —Arnulf Zweig, translator and editor of Kant's *Philosophical Correspondence, 1759-1799*
Paper $18.95 490 pages Library cloth edition $65.00

Distributed for the Center for the Study of Language and Information

Normalization, Cut-Elimination, and the Theory of Proofs
A. M. Ungar

Gentzen's cut-elimination theorem is widely used as a tool for metamathematical investigations, but Ungar investigates the extent to which this theorem and its proofs illuminate the structure of proofs in general.
Paper $18.95 250 pages (est.) Library cloth edition $45.95

Lectures on Linear Logic
A. S. Troelstra

This in-depth volume covers syntactical matters, algebraic semantics, the connection between linear logic and linear categories, proofnets for the multiplicative fragment, and the algorithmic interpretation of cut-elimination in proofnets.
Paper $18.95 225 pages Library cloth edition $45.95

THE UNIVERSITY OF CHICAGO PRESS
5801 S. ELLIS AVE., CHICAGO, IL 60637

Debates and Deviations

Studies in Contemporary German Social Thought

MORAL CONSCIOUSNESS AND COMMUNICATIVE ACTION

Jürgen Habermas

This long-awaited book sets out the implications of Habermas's theory of communicative action for moral theory

244 pp $12.95 paper (November)

CULTURAL-POLITICAL INTERVENTIONS IN THE UNFINISHED PROJECT OF ENLIGHTENMENT

edited by Axel Honneth, Thomas McCarthy, Claus Offe, and Albrecht Wellmer

Focusing on the cultural and political aspects, these essays take up critical theory and political practice, the sociology of political practice, historical-philosophical reflections on culture, and the foundations of critical social theory.

300 pp. $16.95 paper

POSTMETAPHYSICAL THINKING

Philosophical Essays

Jürgen Habermas

224 pp $22.50

MODERNISM AND THE POSTHUMANIST SUBJECT

The Architecture of Hannes Meyer and Ludwig Hilberseimer

K. Michael Hays

"In this original, rigorous, and sophisticated study K. Michael Hays draws on European theory to illuminate the philosophical and epistemological assumptions of the utopian and revolutionary architectural avant-garde of the 1920s and 1930s. Hays breaks new ground in identifying the gradual disappearance of the autonomous subject as a central motif of modern architectural history."

— Anson Rabinbach, The Cooper Union

352 pp, 139 illus. $35.00

NEW PHILOSOPHY OF SOCIAL SCIENCE

Problems of Indeterminacy

James Bohman

280 pp. $32.50

A USER'S GUIDE TO *CAPITALISM AND SCHIZOPHRENIA*

Deviations from Deleuze and Guattari

Brian Massumi

216 pp $12.95 paper

WHAT COMPUTERS STILL CAN'T DO

A Critique of Artificial Reason

Hubert L. Dreyfus

420 pp. $13.95 paper

ZONE BOOKS

ETIENNE-JULES MAREY

A Passion for the Trace

François Dagognet

200 pp, 60 illus. $26.95 (November)

ZONE 6: INCORPORATIONS

edited by Jonathan Crary and Sanford Kwinter

This volume of Zone presents a diverse group of reflections and interventions on the interrelated problems of machine and organism within the context of twentieth-century modernity

600 pp, 200 illus $34.95 paper

Distributed for ZONE BOOKS

THE NEW CONSTELLATION

The Ethical-Political Horizons of Modernity/Postmodernity

Richard J. Bernstein

360 pp $15.95 paper

The MIT Press

From the preeminent publisher of college philosophy texts . . .

INTRODUCTION TO PHILOSOPHY:

Reason & Responsibility:
Readings in Some Basic Problems of Philosophy, 8th Edition
JOEL FEINBERG

The Experience of Philosophy, 2nd Edition
DANIEL KOLAK & RAYMOND MARTIN

Open Questions:
An Introduction to Philosophy, Concise Edition
EMMETT BARCALOW

Writing Papers in Philosophy
ZACHARY SEECH

LOGIC & CRITICAL THINKING:

Logical Reasoning
BRADLEY H. DOWDEN

Open Minds and Everyday Reasoning
ZACHARY SEECH

Arguing About Abortion
LEWIS M. SCHWARTZ

ETHICS & MORAL PROBLEMS:

Basic Moral Philsophy
ROBERT L. HOLMES

What Is The Best Life? An Introduction to Ethics
BRAD ART

Social and Personal Ethics
WILLIAM H. SHAW

Moral Controversies:
Race, Class, and Gender in Applied Ethics
STEVEN JAY GOLD

...Distinguished authors ...and many new titles

FEMINIST PHILOSOPHY AND APPLIED ETHICS:

Feminine and Feminist Ethics
ROSEMARIE TONG

Gender Basics:
Feminist Perspectives on Women and Men
ANNE MINAS

Women and Values:
Readings in Recent Feminist Philosophy, 2nd Edition
MARILYN PEARSALL

Environmental Ethics:
An Introduction to Environmental Philosophy
JOSEPH R. DES JARDINS

Radical Environmentalism:
Philosophy and Tactics
PETER C. LIST

Ethics on the Job: Cases and Strategies
RAYMOND S. PFEIFFER & RALPH P. FORSBERG

AND MORE NEW PHILOSOPHY TITLES:

The Theory of Knowledge:
Classical and Contemporary Readings
LOUIS P. POJMAN

Philosophy of Religion:
An Introduction, 2nd Edition
WILLIAM L. ROWE

An Introduction to Modern Philosophy:
A Core Text
GARRETT THOMSON

For more information on any
philosophy title, contact:

Wadsworth Publishing Company
10 Davis Drive • Belmont, California 94002

Wadsworth presents

A new edition of the most widely used introductory anthology in North America

JOEL FEINBERG'S

Reason and Responsibility:

Readings in Some Basic Problems of Philosophy, 8th Edition

With 16 new selections in this 8th Edition, Joel Feinberg's acclaimed collection keeps getting better. It offers a superb balance of historic selections, recent material, and opposing points of view. Just a sampling of Feinberg's new inclusions:

- Richard Brandt's *Relativism and Ultimate Disagreement About Ethical Principles*
- Jonathan Bennett's "The Conscience of Huckleberry Finn"
- Ayn Rand's "The Ethics of Emergencies"
- Susan Okin's "Justice and Gender"
- Martin Luther King, Jr.'s "Letter from Birmingham Jail"
- Thomas Reid's "Of Identity, with a Criticism of John Locke"
- A. J. Ayer's "Freedom and Necessity"

A new critical reasoning text that's logical, coherent, and fun for your students

BRADLEY H. DOWDEN'S

Logical Reasoning

So interesting that students will enjoy reading it, this book also offers a logical organization that makes it easy to teach.

Argument structure and evaluation are introduced early and Dowden goes beyond teaching students to recognize and criticize arguments by asking them to create arguments for the *other* side—an educational and eye-opening experience!

"Concept Checks" are interspersed throughout for immediate feedback. The book also contains bountiful real-world examples and exercises. A detailed Instructor's Manual is also available.

For more information about our extensive philosophy list, please contact:

Wadsworth Publishing Company
10 Davis Drive • Belmont, California 94002

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 11
NOVEMBER 1992

page
551 *Envy and Inequality* Aaron Ben-Ze'ev
582 *The Identity Theory of Quotation* Corey Washington

606 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

1993

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$65.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$10.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904 Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER. Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 11, NOVEMBER 1992

ENVY AND INEQUALITY*

Envy seems to include the desire to eliminate inequality: the subject's unequal (namely, inferior) position compared to the object's position. In consequence, people raise two different claims regarding the envy-inequality relation: (a) the basis for envy is a concern for equality (and thus a reason for condemning this concern); and (b) reducing inequality will reduce envy (thus a reason for praising the concern for equality). The two claims constitute what seems to be the moral paradox of envy, namely, the fact that envy is morally condemned despite the apparent presence of an important moral component, namely, the desire to eliminate inequality. I believe that both claims are erroneous. I shall argue that envy does not involve a moral egalitarian concern and that reduced inequality does not lead to less envy—on the contrary, in most cases it raises the intensity of envy.

The major question discussed here concerns the dependence of the intensity of envy on the subject-object distance. Is envy more intense when the object's superiority is minor or when it is significant? This question can also be formulated in sociological terms: Will the reduction of social and economic gaps in a certain society lead to a reduction in the intensity of envy? Dealing with these related problems requires the examination of the nature of envy, and in particular whether the focus of concern in envy is the subject's inferiority or the object's undeserved good fortune. In light of the answer to this question, the major variables responsible for the intensity of envy are indicated and the type of correlation between envy and the subject-object inequality is described. In the final sec-

---

* I am grateful to Yoram Amiel, David Heyd, Giora Hon, Ruth Lorand, Saul Smilansky, Daniel Statman, Michael Strauss, and Eddy Zemach for helpful discussions.

0022-362X/92/8911/551–81

tion this correlation is examined by briefly analyzing a society in which inequalities are significantly reduced.

I

Envy may be characterized as a negative attitude toward another person's superiority and the desire to gain what this person possesses. Two major issues appear to be prominent in explaining this attitude: (a) inferiority; and (b) desert. Taking each issue as a central concern of envy results in the formulation of two major approaches: (a) envy is concerned with the *subject's inferiority;* and (b) envy is concerned with the *object's undeserved good fortune.* The first approach assumes the objects of envy to be entitled to their superiority; envy is seen as a nonmoral emotion. In the second approach, envy might be basically considered a justified, moral emotion since the objects are unworthy of their good situation.

Adam Smith's view is compatible with the first approach; he emphasizes the difference in position: "Envy is that passion which views with malignant dislike the superiority of those who are really entitled to all the superiority they possess." This characterization is in accordance with Aristotle's distinction between envy and indignation (or resentment). Both envy and indignation involve pain caused by the sight of other people's good fortune. But whereas indignation is concerned with undeserved good fortune, envy is concerned merely with others' having something that we want. John Rawls takes a similar stand and distinguishes between envy that arises because of "the better situation of others catches our attention," and resentment that addresses injustice and wrongful conduct. Only resentment is a moral emotion. Descartes seems to be close to the second approach and emphasizes the unwarranted position of the object: in envy "we judge the others unworthy of their good." Similarly, A. Ortony and his colleagues regard undeservingness as the central concern of envy and classify envy as a specific example of resentment. General resentment involves a concern for "the undeservingness of the other person enjoying the benefit." Envy adds to this concern the subject's desire to obtain the benefit. Accordingly, they argue that "one's sense that the other does not deserve his good fortune" is of crucial importance in these emotions, and "it is fuelled by one's own sense of privation and by a perception that the other is comparable to oneself."[1]

---

[1] Smith, *The Theory of Moral Sentiments* (Indianapolis: Liberty, 1790/1982), p. 244, Aristotle, *Rhetoric,* II, 9, 10; Rawls, *A Theory of Justice* (New York: Oxford, 1972), p. 533, Descartes, *The Passions of the Soul,* art. 62, in J. Cottingham, R. Stoothoff, and D. Murdoch, trans., *The Philosophical Writings of Descartes* (New York: Cambridge, 1984); Ortony, G. L. Clore, and A. Collings, *The Cognitive Structure of Emotions* (New York: Cambridge, 1988), pp. 101–2.

The assumption that moral concern is central to envy is fraught with difficulties. To begin with, envy is often directed at people who are lucky, but being lucky is not immoral. Moreover, envy addresses a partial concern rather than the general concern typical of the moral domain. It entails the desire to improve the subject's personal lot, not concern for other people. This is in accordance with the discriminatory and partial nature of emotions in general (to be discussed below). Furthermore, sometimes the envious person wishes to deprive the object of her greater benefits, even if this means depriving oneself of some benefits as well. There can hardly be moral justification for such a desire; it would make sense only if the comparative inferiority of the subject were of crucial importance.[2]

If moral concern is not essential to envy, its classification as a particular example of resentment is inadequate. Resentment may be characterized as an emotional protest against what is perceived as morally unjust. Accordingly, it is closer to anger than to envy. Although resentment is sometimes directed at the superiority of others, the focus of the subject's concern is some kind of moral attitude rather than superiority itself. Here we may distinguish between discontent, envy, and resentment. Discontent arises when we believe that something (of some importance to us) is wrong, namely, a better alternative is available. Envy occurs when the wrongness is related to our inferior situation; resentment occurs when wrongdoing is perceived: it conveys an implicit accusation. When unjust treatment, but not inferiority, is perceived, anger and resentment are more dominant than envy. The relation between envy and resentment is somewhat similar to that between regret and remorse, except that the latter are directed at oneself. Like resentment, remorse is usually directed at moral wrongdoing; as in envy, the main concern in regret is undesirability or dissatisfaction and not necessarily immorality. Furthermore, resentment and remorse presuppose some responsibility of an agent (object in resentment and subject in remorse) for the immoral behavior; such responsibility may be absent from envy and regret.[3] Envious people often like to em-

[2] The readiness to suffer in order to make the other person suffer more is illustrated in the story about the angel who came to earth and promised a farmer to fulfill any wish, though on one condition. his neighbor would receive twice as much of the same. In that case, said the farmer, please remove one of my eyes. On the relationship between moral and emotional attitudes, see J. Oakley, *Morality and the Emotions* (New York: Routledge, 1991); my "Emotional and Moral Evaluations," *Metaphilosophy*, XXIII (1992): 214–29

[3] See R. Folger, "Perceived Injustice, Referent Cognitions, and the Concept of Comparison Level," *Representative Research in Social Psychology*, XIV (1984). 88–108; W. G. Parrott, "The Emotional Experience of Envy and Jealousy," in P.

phasize their concern for moral justice, thus attempting to justify it. Accordingly, they tend to describe their attitude as resentment rather than envy. It is clear, however, that this is often a kind of rationalization of their negative attitude to being inferior. Hence, the transformation of envy into resentment is common, whereas the opposite transformation is rare.

If desert is not the central concern of envy, the natural candidate for such concern is inferiority. The importance of the inferiority concern in envy conveys the weight we attach to our comparative stand. People compare themselves with others in order to reduce uncertainty about themselves and maintain or enhance self-esteem. An unfavorable comparison often leads to envy. Social comparison appears in many forms; an obvious form typical of envy is competitiveness. Our inferiority displeases us since it is perceived as reflecting some loss in a certain ongoing competition. The differences among different people are often perceived as reflecting different positions in an underlying competition. The connection between competitiveness and envy is aptly described by A. Kohn:[4]

> Competition creates a prized status where none existed before, thereby giving us something to desire. Then it insures that not everyone can get it. Finally, competition requires that those who obtain the reward can do so only by defeating everyone else. Both the objective and subjective conditions for envy are established, in other words: restricted access to something desired and a (quite accurate) belief that someone else has got it at one's own expense (*ibid.*, p. 141).

Since we usually do not compete with dead people, we usually do not envy them and hence can afford to be generous. Following Hobbes, we may even say that praise of ancient authors frequently proceeds from envy of the living with whom we compete rather than from reverence of the dead. When we consider ourselves as competing with the dead, as when we want to outstrip their achievements, we may envy or be bitter toward them. Aristotle[5] argues that the lack of competition is also the reason why we do not envy those not yet born. It should be noted that envy presupposes a certain social com-

---

Salovey, ed., *The Psychology of Jealousy and Envy* (New York: Guilford, 1991), pp. 8–9, G Taylor, *Pride, Shame and Guilt* (New York: Oxford, 1985), p. 98.

[4] *No Contest: The Case Against Competition* (Boston: Houghton Mifflin, 1986)

[5] *Rhetoric*, II, 10 It seems, however, that we envy future more than past generations. One reason is the constant improvement in the quality of life. (Since such rapid improvement was less typical in ancient times—it was even assumed that the golden age was over—this factor has lesser weight in Aristotle's view.) Another reason is rooted in the fact that it is easier for us to imagine a past world without us, than a future world.

parison, but not necessarily competition (though it typically does). Competition is one form of social comparison. My wish to be as wealthy as Rockefeller involves social comparison, but not competition. I cannot conceive of myself as competing with Rockefeller in this regard.

It is our relative deprivation, rather than all types of deprivation, which bothers us in envy. A rich person can be envious of another having a bit more; a beautiful movie star can be envious of another star's beauty. "Enviers want to be better more than they want to be better off"; hence, "kings may seem to have almost everything good a person could have, but they may envy if there is anything good they yet lack and if they think that other kings have some advantage over them."[6] The notion of relative deprivation implies a distinction between the objective and subjective well-being of the subject, and the claim that the two are not isomorphically related; sometimes the better off from an "objective" viewpoint feels subjectively worse off. Thus, a survey reveals that first-line supervisors earning between $12,000 and $15,000 a year were more satisfied with their salaries than company presidents earning about $40,000 a year. It is often true that "those who are the most deprived in an objective sense are not the ones most likely to experience deprivation."[7] The fact that envy involves relative deprivation emphasizes the importance of the inferiority concern in envy.

It may be argued that envy emerges not only when social comparison indicates the subject's inferiority, but whenever we desire something that we lack; for example, when we are tired or wish to have more money. I believe, however, that such cases represent discontent or covetousness rather than envy. In discontent and covetousness, social comparison is not an essential component as is the case in envy. Unlike covetousness and discontent, which are merely concerned with gaining something or achieving a certain state, envy is mainly concerned with someone else who has something or is in a certain state. I covet money and am discontented with my financial situation, but I envy the rich. Unlike envy, covetousness does not involve a belief in the subject's inferiority vis-à-vis the person who has the desired object. In fact, people of superior standing often covet what belongs to people in an inferior position. Covetousness is more similar to greed than to envy. Like greed, it does not cease when the desire is achieved, and like kleptomania it is directed at something one does not really need.

---

[6] Montaldi, "Envy" (unpublished manuscript). It may be argued that envy does not represent the wish to eliminate inequality but to achieve superiority. Although I agree that the latter wish may be found in some cases of envy, it is the former which is more essential to envy. The latter wish is often found when envy is not present and many cases of envy lack it.

[7] F. Crosby, "A Model of Egoistical Relative Deprivation," *Psychological Review*, LXXXIII (1976): 85–113, p. 85 See also Kohn, J. Neu, "Jealous Thoughts," in A O Rorty, ed., *Explaining Emotions* (Berkeley California UP, 1980). The relative deprivation typical of envy may, however, be connected with what the envious person perceives as nonrelative deprivation of respect or attention that the subject would like to gain. Indeed, the quest for respect or fame is a crucial element of envy

To conclude the first stage of our discussion: it is quite clear that the central element of envy is not a moral concern but perceived inferiority. If inferiority were the only concern in envy, we should expect a positive correlation between the intensity of envy and the subject-object inequality. In the next section I shall show, however, that this is not the case and that the relation between envy and inferiority is complex.

II

Unlike the proposed correlation between envy and inferiority, envy seems to be directed at those who are like us or equal to us but are still slightly superior to us. The Greek poet Hesiod already wrote that "potter is furious with potter and craftsman with craftsman, and beggar is envious of beggar and singer of singer." Aristotle argues that "we envy those who are near us in time, place, age, or reputation." Similarly, Bacon observed that "where there is no comparison, no envy; and therefore kings are not envied but by kings."[8] In envy, our attention is focused on those perceived to be immediately above us, since (as G. Simmel remarks) these people occupy the first rungs on fortune's ladder that we will have to climb. These are the people we are most likely to be compared with or whose accomplishments are most likely to demean us. This has been described as "neighborhood envy": each person in a hierarchy primarily envies the person immediately above himself. In cases of extreme inequality, especially cases of unattainability, far less envy is aroused than in cases of minimal inequality, which inevitably provokes the envious one to think, "I could be in his place." Although in typical envy we imagine ourselves in the object's place, we do not lose our identity; we keep it while imagining an improvement in our situation. It is easier to keep our identity while imagining ourselves in the place of those similar to us.[9] A woman may envy another woman for her beauty, but she will probably not envy a handsome man. Where no proximity exists, comparison is less likely to arise and we are less prone to feel inferior. Those who are close to us, but still above us, emphasize our own inferiority more than those who are distant from us. Moreover, the issue of desert is more central in closer proximity: the subject-object comparison is easier, the subject's (slight) inferiority is not taken for granted and its presence is more likely to be perceived as unfair.

---

[8] See P. Walcot, *Envy and the Greeks* (Warminster. Aris & Phillips, 1978), p. 9; Aristotle, *Rhetoric*, II, 10; Bacon, *Essays*, in *Selected Writings* (New York: Modern Library, 1625/1955), p 9 Kant's position on this issue is interesting Although he argues that "Persons of the same station and occupation in life are particularly prone to be jealous of each other," he contends that people compare themselves with others in order to achieve a better self-image. A favorable comparison is likely to emerge since people "choose as a rule the worst and not the best of the class with which they set up comparison; in this way their own excellence shines out." See Kant, "Jealousy and its Offspring—Envy and Grudge," in *Lectures on Ethics* (New York. Harper, 1963). Kant's position is consistent if we assume that we choose similar people as constituting our comparative framework, but within this framework the particular individual chosen for comparison is that occupying a somewhat inferior position Although it is true that in order to cheer up we sometimes compare ourselves to those inferior to us, such an escape device is often not helpful; otherwise, envy would not be so common.

Research concerning levels of aspiration confirms the prominence of the proximity variable. It has been shown that in most situations the level of aspiration is placed slightly above the subject's performance. This is in accordance with the tendency of the level of aspiration to stay out of too difficult or too easy areas. It is also in accordance with our wish to be slightly better than those with whom we compare ourselves. An interesting study was made of people with different salaries. They were asked how much of a raise they required in order to live comfortably. Most subjects in all salary ranges indicated that an increase of about 15 percent would suffice. To put it crudely, this is usually the range of our envy. Empirical evidence suggests that competitive comparison is maximal in small subject-object gaps in current positions, that is, when the subject is slightly inferior to the object.[10] The factor of the availability of an alternative, which is a major factor in the generation of envy and other emotions, is most significant in such cases (see below).

The subject-object proximity typical of envy and other emotions is compatible with another salient characteristic of emotions: their partial and discriminatory nature. Not everyone and not everything is of emotional significance to us. We usually would not assume an emotional attitude toward someone completely unrelated to us. Thus, when we hear of the death of thousands of people in an earthquake occurring in a remote (that is, from our vantage point) part of the world, our emotional response comes nowhere near the intensity of our grief at the death of someone close to us, nor does it approach the level of feeling we experience in watching the death throes of a single victim of that same earthquake on television (thereby establishing some proximity with this particular victim). Emotions require many resources such as time and attention. Since these resources are limited, emotions must be partial and discriminative.[11] Subject-object proximity frequently employs a partial and momentary perspective that contradicts the impartial and broad perspective typical of scientific knowledge. When we look at someone from a close distance, our vision of this person is often distorted. In the extreme case when there is no distance at all, namely, when one places the object upon the eye itself, one does not see the object. We need some distance in order to achieve a perspective that encompasses multiple aspects of the object and thereby makes the object less fragmentary. But keeping a distance is contrary to the involved and interested perspective typical of emotions. Accordingly, the broader the perspective, the lesser the intensity of emotions. Thus, a perspective that does not merely focus on the object's superior position but also on that person's character may reduce envy. If the object is kind and modest, envy may decrease since it can be assumed that the object respects the subject.[12] Envy may also decrease if the object is mean since the negative evaluation of the object's character will outbalance the object's superiority (accordingly, it has been claimed that "only the shameless envy the wicked"). What matters in reducing envy in these cases is not so much how we evaluate the object, but that we do so from a broader perspective referring to many aspects of the object. The broader the perspective, the lesser the weight of the particular inferiority that generates envy.

---

[9] Cf. J. Elster, "Envy in Social Life," manuscript (1991), H Schoeck, *Envy* (New York: Harcourt, 1970), pp. 62, 237; M. Silver and J. Sabini, "The Perception of Envy," *Social Psychology*, XLI (1978): 105–17, D Heyd, *Genethics: Moral Issues in the Creation of People* (Berkeley: California UP, 1991), p. 187

[10] See L. Festinger, "A Theory of Social Comparison Processes," *Human Relations*, VII (1954) 117–40; K. Lewin, T. Dembo, L Festinger, and P. S. Sears, "Level of Aspiration," in J M Hunt, ed , *Personality and the Behavior Disorders, Vol. 1* (New York: Ronald, 1944), pp. 333–78; S Dakin and A. J. Arrowood, "The Social Comparison of Ability," *Human Relations*, XXXIV (1981) 89–109. The research concerning the different salaries was conducted over forty years ago I assume that by now the range of our envy is wider because envy is stronger in contemporary society (see below).

---

[11] Spinoza argues that "the affects are excessive, and occupy the mind in the consideration of only one object so much that it cannot think of others." For example, "a greedy man thinks of nothing else but profit, or money, and an ambitious man of esteem" (*Ethics*, IV, Prop. 44, Schol.).

[12] Isocrates, the ancient Athenian orator, says that we envy those who surpass us in intelligence or in any other respect, unless we are won over by their daily kindness. See Walcot, p. 35. The absence of envy in this case may be due to our interpretation that these people deserve their good fortune (indicating that desert has some role in envy) or that their kindness implies respect for others and hence they lack a feeling of superiority. (Since envy is a complex state, it may be also directed, because of other reasons, at virtuous, superior people. If we could isolate the effect of being virtuous, however, the intensity of envy would be reduced in most cases.)

In light of the partial nature of envy, the relevant inferiority concern is also partial. It addresses people who belong to our emotional environment, that is, those of emotional significance for us. The inferiority concern in envy is partial (or specific) in two important senses: (a) the subject's position is low compared to a *specific person:* it need not be low in comparison with the whole population, the average position in society, or even the top five percent; (b) the subject's position is low in a *specific domain* particularly relevant to the subject's self-evaluation. Social comparison is not exercised indiscriminately, but refers to people and domains relevant to our self-evaluation. We compare ourselves with people whom we consider to occupy an approximately similar position or possess similar ability. We tend to exclude from our reference group people who appear definitely superior or inferior to us as well as those belonging to irrelevant domains.

The specific nature of the inferiority concern in envy is related to the subject's concern with relative and not general deprivation. Envy implies a particular situation worse than that of someone else of importance to us. We may envy only one aspect of another person, continuing to consider ourselves superior in general. We may envy someone for being rich, but if we know this person to be dishonest, we consider ourselves superior. Generally, a subject, clearly superior to an object, may be envious of the same object who surpassed her in a specific sphere. In such cases, envy would not usually be intense. Similarly, we may desire to possess something, but if this thing does not belong to our self-defining domains, envy will not arise or will not be intense. The lack of these things does not make us feel inferior since we "either have enough other good things to which our self-esteem is linked, or we have found deeper satisfaction elsewhere."[13] Whether we admire or envy another person is determined, among other things, by the relevance of that person's good fortune to our self-evaluation. When the good fortune is relevant, our self-evaluation and our evaluation by others is threatened and envy arises. This is particularly true of domains in which large discrepancies exist between actual and ideal self-description. Here the threat to the subject's self-esteem is especially strong. This kind of specificity distinguishes typical adult envy from envy typical of children. General envy, nonspecific to certain people or domains, is typical only of young children for whom the value of most things comes simply from seeing others with them. Young children, who lack a well-defined sense of what is relevant or irrelevant to themselves, feel discomfort every time they fail to measure up to another's performance. Older children are more selective in this respect: failure in domains of importance elicits more envy than in marginal domains. The specific nature of the inferiority concern in adult envy explains why, contrary to common belief, psychological research has failed to reveal a significant positive correlation between envy and an individual's general low self-esteem.[14]

---

[13] Montaldi; my "Envy and Jealousy," *Canadian Journal of Philosophy*, XX (1990): 487–516

The foregoing discussion has indicated that envy is concerned with specific inferiority, not inferiority in general. This conclusion is incompatible with the assumption that inferiority is the sole concern in envy and that we should therefore expect a positive correlation between the intensity of envy and the subject-object gap in current positions. Since social comparison is chiefly limited to those similar to us, envy should be more typical of small subject-object gaps. Moreover, in small gaps the question of desert is quite prominent. In comparison to people similar to us, we would presumably consider our inferior situation incompatible with what we deserve. Indeed, emotional intensity is greater in unjustified than in justified circumstances. This is because the situation is considered to be more abnormal and the availability of an alternative is greater. Furthermore, if a strong wish to obtain something that another person has were to persist for a long time, the subject would tend to feel in some degree entitled to the desired object. Desert has after all some role in the generation of envy. Let us, therefore, examine the issue of desert.

III

Both the approach considering inferiority to be the sole concern in envy and the approach characterizing envy as a moral emotion assume that desert claims are to be identified with moral claims. Accordingly, the second approach, which realizes that desert has some role in the generation of envy, identifies envy with resentment and considers both to be moral emotions. The first approach, realizing that in many cases the moral concern is absent from envy, concludes that desert has no role in the generation of envy. The identification of desert claims with moral claims is explicitly stated by a few philosophers. Thus, Aristotle argues that "whatever is undeserved is unjust," and J. Hospers claims that "justice is (simply) getting what one deserves."[15] I believe that the most plausible way to explain envy and resolve the moral paradox of envy is to reject this assumption and distinguish between moral and desert claims.

---

[14] S. A. Bers and J. Rodin, "Social-Comparison Jealousy. A Developmental and Motivational Study," *Journal of Personality and Social Psychology*, XLVII (1984): 766–79; P. Salovey and J. Rodin, "Envy and Jealousy in Close Relationship," in C. Hendrick, ed., *Close Relationships* (Newbury Park: Sage, 1989); A. Tesser and J. Campbell, "Self-Definition: The Impact of the Relative Performance and Similarity of Others," *Social Psychology Quarterly*, XLIII (1980) 341–7.

The distinction between moral and desert claims is not easy to make and its detailed description is beyond the scope of this discussion. Nevertheless, I would like here to point out some important characteristics of the distinction which are relevant to the role of desert claims in envy. Whereas desert claims are based on the value of a person's attributes and actions, moral claims often refer to obligations toward other persons. Desert claims are not necessarily grounded in anyone's obligations, but rather, as G. Sher[16] rightly claims, "in the value of persons' coming to have what they deserve." This is so, since if "persons themselves matter, then what matters to persons should matter as well." Because of the more private nature of desert claims, their validity is often grounded in preinstitutional values; desert, Sher argues, "preserves its normative independence from institutional arrangements." A major reason for the private nature of desert claims is that they are often based on one's personal desires.[17] In perceiving our situation to be undeserved, we do not necessarily accuse a certain person of behaving criminally or immorally; we assume, however, that our being in such a situation is in some sense unfair. Whereas in moral claims the agent is a person having some responsibility, an impersonal cause can also be an agent in desert claims. Heavy rain may be the cause of undeserved but not of immoral circumstances. Similarly, whereas moral claims are typically directed at a human being (and sometimes also other creatures), desert claims may also be directed at inanimate object. One can say (as Sher notes) that "Cleveland deserves better publicity, since it is an interesting city." Desert claims are based on *perceived* undeserved situations that are not necessarily undeserved in a more objective sense. Perceived undeserved situations may be due to impersonal, arbitrary circumstances. Being unlucky may not involve any criminal or immoral deeds and attitudes of an agent, but the unlucky person may still rightly regard it as undeserved. It is not immoral for a rich person to win a big prize in the lottery or to marry a rich person; nevertheless, many poor people may consider it to be undeserved and envy this person.

---

[15] Aristotle, *Rhetoric*, II, 9. Hospers, *Human Conduct: Problems of Ethics* (New York: Harcourt Brace Jovanovich, 1972), p. 361.

[16] *Desert* (Princeton: University Press, 1987), pp. 195, 58, 205 My discussion of desert is heavily indebted to Sher's excellent analysis. Whereas Sher considers "desert" a moral notion that is distinct from "right" and "justice," I use "desert" in a broader sense and allow for desert claims that do not belong to the moral domain

[17] This is in accordance with the claim that desires confer value on their objects. Spinoza, e.g., argues that "we judge something to be good because we strive for it, will it, want it, and desire it" (*Ethics*, III, 9, Schol.) For further discussion of this matter, see Sher, p. 55. See also D. A. Prentice and F. Crosby, "The Importance of Context for Assessing Deservingness," in J. C Masters and W. P. Smith, eds., *Social Comparison, Social Justice, and Relative Deprivation* (Hillsdale: Erlbaum, 1987), pp. 165–82.

The desert claims typical of envy are personal and only rarely can be satisfied by a moral action. Such claims are not considered as serious moral claims. Concerning many of them, telling the subject to be satisfied with his own fortune can be considered to be a proper moral response. Moreover, in most cases there is no one to whom to address desert claims, and anyhow fulfilling these claims cannot be done in light of practical considerations. When David envies Adam's beauty, his envy involves a desert claim but not a serious moral claim. Telling David that he should be satisfied with his own good fortune in other domains is in this case a proper moral response. There is no one to blame for Adam's beauty and it is impractical to change the situation by doing a plastic surgery on David and all other people who are not as beautiful as Adam.

A distinction can be made between deserving and wishing. When we express our wish to have something, we do not necessarily imply that our lack of this thing involves unfairness or injustice. In claiming that we deserve something, we do suggest there is something unfair in the fact that we do not have this thing. Envy involves not merely a wish claim but a desert claim as well. Envy is often based on personal, nonmoral norms of desert, whereas resentment is usually based on societal moral norms of justice.[18] Although envy typically involves desert claims, it may sometimes involve moral claims. Envious people often describe the unfairness involved in their situation as moral; they thus attempt to legitimize their envy and may even present it as resentment rather than envy. Since the borderline between the moral and desert claims is not clear-cut, it is not always easy to distinguish between envy and resentment.

---

[18] The urge to find some kind of unfairness in the subject's inferior position may also be explained by referring to the saying that "Injustice is relatively easy to bear; it is justice that hurts." A similar view to the one presented here is found in Richard Smith, "Envy and the Sense of Injustice," in P. Salovey, ed., *The Psychology of Jealousy and Envy* (New York: Guilford, 1991). Smith speaks about "the subjective, unsanctioned nature of the sense of injustice in envy." He rightly argues that "In an unconventional but arguable sense, people on the short end of the distribution have been unfairly treated by fate." Accordingly, "the envied person's advantage is, to some degree and on some subjective level, unfair" (pp. 85, 92–3). This valuable article clearly shows the need to distinguish the sense of injustice typical of envy from that typical of resentment. Smith refers to the unfairness typical of envy mainly in order to explain the element of hostility in envy. I think that this sense of unfairness principally explains the element of discontent in envy. Since this sense is often not directed at culpable deeds or attitudes of a specific agent, but at impersonal circumstances, its major explana-

It should be clear by now that the issue of desert plays some role in the generation of envy. This is not, however, the role suggested by the aforementioned two approaches. The first approach is right in arguing that inferiority is of concern in envy, and the second is right in assuming that desert is of concern in envy. The first is wrong in believing that this concern refers to what the subject considers to be deserved inferiority, and the second is wrong in claiming that the concern of desert is directed at the object. Accordingly, the focus of concern in envy is not merely the ***subject's (deserved) inferiority***, nor the ***object's undeserved good fortune***. Envy is rather concerned with the ***subject's undeserved inferiority***. The issue of desert is always accompanied by the issue of the subject's particular inferiority. The subject's concern is not a general moral concern for justice, but a particular personal concern for what the subject considers to be undeserved inferiority. Sometimes envy includes the suspicion that the subject's inferior position indeed reflects what the subject actually is. Even in this case the subject may still feel the situation to be unfair since being inferior is unfair. Here the complaint is not against the wrongdoing of people but against a divine or impersonal cause.[19]

The element of desert is not obvious in some cases of what is commonly characterized as envy, for example, when I am hungry and see someone eating. The issue of desert is clearly present if I do not have the money to buy food. If I do have the money, however,

---

tory role does not concern personal hostility toward a certain agent but a more general feeling of discontent.

[19] In many respects pleasure-in-others'-misfortune (*Schadenfreude*) is similar to envy. Like envy, typical cases of pleasure-in-others'-misfortune involve small differences and desert claims. Thus, this emotion may be characterized as delight in someone's deserved and usually minor misfortune. When the object's situation becomes highly undesirable, pleasure-in-others'-misfortune often turns into pity. Pity is concerned with substantial suffering and not with minor misfortunes. Pleasure-in-others'-misfortune conveys the belief that some justice has been done. The more just the misfortune is, the better justified is the pleasure in it. Indeed, we often hear people who delight in others' misfortune saying that "God has punished this person," or that "there is justice in the world." See my papers, "Pleasure-In-Others'-Misfortune," *Iyyun*, XLI (1992): 41–61; and "Another Look at Pleasure-In-Others'-Misfortune," *Iyyun*, XLII (1993).

but do not enter the restaurant because I am in a hurry, the issue of desert is hardly relevant. But such a case should not be regarded as envy—anyway not a typical case of envy. It concerns a momentary deprivation and not relative inferiority. Moreover, if I am in a hurry to do something that I like, I value this thing more; accordingly, the momentary hunger will not be of significance and it may be overcome by eating a sandwich. If I am hungry because I am obliged to attend a certain tedious meeting, I may think that there is something unfair in my situation; in this case my attitude will be closer to envy. Deprivation in satisfying our needs is not identical with envy. Envy involves the subject's relative inferiority and the belief that this deprivation is undeserved. Consider another example. Moses considers himself lazy but nevertheless envies Dean's success, which is the result of Dean's hard work. If Moses believes that his laziness is due to an innate disposition that he cannot overcome, he may consider his situation unfair, and envy is likely to emerge. If Moses can overcome his laziness, however, but chooses not to do so as he considers Dean's success to be not worth the price of not being lazy, then envy is not likely to appear, or it will at least not be intense. Moses may want the results of Dean's success, but since typical envy is directed at a person and not at a certain thing that the person has, this would not be typical envy in which the subject is ready to exchange his present position with that of the object.

The relation between desert and the intensity of envy is positively correlated: the more we believe our inferior situation to be undeserved, the greater the intensity of envy. Indeed, there is some empirical support to the claim that the tendency to feel inferior to others and to construe one's disadvantages as due to unfairness predisposes a person to be envious.[20] When we believe that we deserve an inferior position, for example, when the object is perceived to be much better than us, the competitive concern is hardly present and envy usually does not arise.

We are now in a position to resolve the moral paradox of envy arising from the fact that it is morally condemned despite its apparent moral value embodied in the desire to eliminate inequality. The paradox is resolved by referring to the combination of the two major concerns in envy: desert and inferiority. The perceived undeservingness in envy does not express moral injustice; hence, envy is not a moral emotion. Moreover, although we consider our situation to be unfair, to express it in public means to admit our inferiority and this is difficult to do. Nevertheless, envy involves value claims concerning the situations in which we think we deserve to be. Hence, it has some normative justification.

---

[20] See Parrott.

Returning to our original question concerning the correlation between the intensity of envy and the subject-object inequality, the answer would be simpler if envy could be defined in terms of mere inferiority or desert. If envy were related to inferiority in general, it would be plausible to assume that the greater the inequality, the more intense envy would be. If envy were related merely to desert, it would be plausible to postulate a mainly negative correlation between the intensity of envy and the subject-object inequality, because questions of desert are more prominent in smaller inequality, where inferiority is less evident. Since we define envy as concerned with undeserved inferiority, it is not immediately obvious what the relationship is between the intensity of envy and the subject-object inequality. Take, for example, the case in which our colleague wins a professional prize (which presumably reflects the recipient's merits). If the colleague indeed deserves the prize, then our inferiority is evident and envy is likely to arise. If the colleague does not deserve the prize, we may feel that we should have received it, once again provoking envy. The first case is more typical of envy, the second of resentment. In both cases inferiority as well as desert play a role in generating the negative emotion. In the next section, I shall try to clarify the complex relationship between the intensity of envy and the subject-object inequality.

IV

The notion of "emotional intensity" is quite complex and cannot be discussed at length here. I would merely like to point out a few major considerations that are relevant for our discussion. Emotions occur when we confront significant changes highly relevant to our own life. When we experience an emotion, our physiological and mental systems cease to maintain their usually stable state. Emotions indicate a transition in which a preceding context has changed, but no new context has yet stabilized. The unstable and transient nature of emotions is due to significant changes, exceedingly relevant to the subject's situation (and the situation of those related to the subject). The stronger, or the more real the change, the more intense the emotion. Accordingly, the basic parameter determining emotional intensity is the one that measures the impact of the change upon the subject's situation. As this parameter measures the reality of the change in the subject's eyes, it may be termed a "sense of alteration."[21]

---

[21] Spinoza strongly emphasizes the importance of changes in the subject's situation for the generation and intensity of what he terms "passive emotions." He

The "sense of alteration" involves two kinds of change in the subject's situation: (a) perceived actual changes, and (b) imaginary possible changes. The first kind refers to the perception of factual changes that have taken place, while the second refers to imagined counterfactual changes. The first kind of change refers to actual modifications in the subject's situation. The greater this change, the more intense the emotion. The more the subject believes the emotion-inducing situation to be real, the more intense is the emotion. When we know that the danger is real, we are more frightened than when we think that the danger could be illusory. Likewise, we shall be less envious in seeing a successful person in a movie than in a real situation. The second kind of change typical of emotions is imaginary. It encompasses a mental construction of a counterfactual alternative, to be compared with actual reality. Because of the significance of the change for the subject, the subject is keenly aware of an alternative to the current situation. The more available the alternative is, namely, the closer the imagined alternative is to reality, the more intense the emotion is. Availability (or feasibility) is determined by how remote from actual reality the relevant event is considered to be. Accordingly, Spinoza argues that the better the wise man grasps the fully deterministic nature of events around him, the less likely he is to experience passions. That is, the better a person realizes the lack of alternatives, the less likely it is for this person to be affected by emotions. Generally, "insofar as the mind understands all things as necessary, it has a greater power over the affects, or is less acted on by them." Thus, Spinoza[22] argues that no one pities an infant because it cannot speak, walk, or reason; but if most people were born full-grown and only one here and there as an infant, everyone would pity infants.

Although the first kind of change constituting the sense of alteration may be regarded as more objective than the second, it is also

---

claims that "we live in continuous change, and . . . as we change for the better or worse, we are called happy or unhappy" (*Ethics*, v, Prop. 39, Schol). Among contemporary philosophers, W. Lyons most strongly emphasizes the role of change in generating emotions. He argues that a subject is in an emotional state if and only if he is in an abnormal physiological state caused by his evaluation of the context in relation to himself; see *Emotion* (New York: Cambridge, 1980), ch 7 Similarly, the psychologist R. Lazarus argues that an emotional state "is a transient reaction to specific encounters with the environment, one that comes and goes depending on particular conditions"; *Emotion and Adaptation* (New York: Oxford, 1991), p. 47. He distinguishes an emotional state from an emotional trait that "is not an emotion but a disposition to react with one" (p. 47). For further discussions dealing with some of the difficulties concerning the role of change in generating emotions, cf. my *Emotions toward Others* (forthcoming).

[22] *Ethics*, v, Prop 6

essentially a subjective change. It is the subject who perceives the change and accordingly considers it more or less significant. Moreover, changes associated with emotions may not merely refer to the subject but also to those constituting the subject's environment. Again, it is the subject who determines which people belong to this environment. The subject not only determines the significance of the change but also its scope. In a way the two kinds of change are similar in measuring the degree of reality of the given change in the subject's situation. We construct a psychological reality in which some changes are smaller than others, since they have less influence on the subject's situation, and some alternatives are closer to reality than others, since they are more likely to happen. In constructing our reality, we assume that some modifications are much more natural and feasible than others, and some attributes are particularly resistant to change.[23]

The "sense of alteration" is more significant in small subject-object gaps where every change can significantly alter the relative position of the subject. For example, changes from a superior to an inferior position and vice versa are very probable in such gaps. When the subject-object gap is large, most changes in the subject's position will be insignificant for the subject's relative inferiority and hence will hardly have emotional significance. This is true concerning both actual changes and imaginary changes. To illustrate this point, consider the following example by Kahneman and A. Tversky:[24]

> Mr. C. and Mr. D. were scheduled to leave the airport on different flights, at the same time. They traveled from town in the same limousine, were caught in a traffic jam, and arrived at the airport 30 minutes after the scheduled departure time of their flights. Mr. C. is told that his flight left on time. Mr. D. is told that his flight was delayed, and just left five minutes ago. Who is more disappointed?

Mr. D. is more disappointed since catching up with his flight was almost possible for him. Mr. D.'s flight was only five minutes away from actual reality, whereas Mr. C.'s flight was half an hour away. It is easier for Mr. D. to construct a reality in which he catches up with his flight than it is for Mr. C. Disappointment is stronger when one can imagine oneself as being close to the desired object. Similarly, we envy people when we can easily imagine ourselves in their place; hence, we envy more those who are like us or equal to us. In light of the importance of the availability of the alternative, "almost situations" have additional emotional nuances. As F. Heider[25] argues, if we hold number 5304 in a lottery and we learn that 5305 is the winning number, this near success is probably harder to bear than if there were no winning number anywhere near our town. Generally, a near success is more frustrating than a simple failure, and misfortune appears less negative when it is thought of as necessary than when it is seen as something that could easily have been prevented.

---

[23] See D. Kahneman and D. T. Miller, "Norm Theory: Comparing Reality to its Alternatives," *Psychological Review,* XCIII (1986): 136–53, pp. 142–3 These authors claim that "A counterfactual possibility should appear 'close' if it can be reached by altering some mutable features of reality."

[24] "The Simulation Heuristic," in Kahneman, P. Slovic, and Tversky, *Judgment under Uncertainty: Heuristics and Biases* (New York: Cambridge, 1982); Kahnemann and Miller, "Norm Theory."

The subject's inferiority is more obvious but less significant in large gaps. In great subject-object gaps, when subjects find it impracticable to establish equality, they no longer find this inequality disturbing. Our emotional sensitivity to others' privileges, and hence our own inferiority, increases as privileges become rarer and more controllable; attaining the alternative is easier in these circumstances. Accordingly, slight inequality has often greater emotional impact than great inequality.

Assuming a necessary linear correlation between the intensity of envy and the subject-object distance is then a false assumption which ignores the complex nature of envy. Envy may occur in slight and very great proximity—though it usually will not be intense in these circumstances. Envy may also not occur in great similarity of background such as is found in very close relationships. A Hebrew proverb states that a person is never envious of a son or a pupil. (Some of us doubt if this is even remotely true concerning one's pupil.) Achievements of those very close to us evoke pride rather than envy when these achievements are understood (by us or by society) to be connected to us in such a manner that we can share the credit they bestow. Moreover, very close relationships are complex and rich; hence, it is not likely that envy, which focuses upon a partial aspect, is common in these relationships. A father usually considers his son's success as part of his own success and not as something separate that threatens his own self-esteem. When those close to us have succeeded in something essential for our own self-evaluation, however, or when our ties with these people are not evident, envy may replace pride. Thus, envy can play a part in some close relationships such as that between parents and children, siblings, or husband and wife. A father may wish to be at his son's age, to have his son's sexual or economic opportunities, etc. He may also perceive his own situation to be inferior to that of his son. But a father's attitude toward his son rarely becomes that of intense or malicious envy. For example, a father will usually not aspire to damage his son's position in order to nullify the inequality between them. If such intense envy does occur, we can assume that the father-son relationship is not very close. It seems that it is more common to see children who envy their famous parents than parents who envy their famous children. This shows that parents are closer to their children than vice versa.[26] Generally, social comparison underlying envy requires not only a certain proximity, which enables us to place the two items in the same comparative framework, but also a certain distance that enables us to see the two items as distinct.

---

[25] *The Psychology of Interpersonal Relations* (New York: Wiley, 1958), pp. 141–4. Needless to say, the generation of envy and other emotions is also a function of one's personality and prevailing social norms. Thus, someone may be less envious by nature than other people, or a certain society may discourage envy. It is assumed here that these factors are given and the task is to find the variables responsible for generating envy. In some cases personality (and to a lesser extent social) factors are changed as the result of the emotional experience, thus becoming effects rather than causal antecedents of an emotion (Lazarus, p. 204).

Sometimes envy is also present in cases of large subject-object gaps. Thus, we may envy rich or famous people with whom we hardly compare and in any case do not think of competing. The existence of such cases does not contradict our contention that envy is typical of small subject-object gaps. Typical is not the same as universal. The tendency to compete with a noncomparable agent is reduced but does not disappear in small gaps because conditions of noncomparability are relative and continuous, rather than a case of all or nothing.[27] Furthermore, the subject-object gap is not the sole variable in determining the generation of envy; other variables may sometimes be effective even with a large subject-object gap. Moreover, the inferiority in large gaps is more profound, thought to be inadmissible, and cannot be ignored. Another aspect complicating the situation is that envy, like other emotions, does not appear in isolation but as part of a complex emotional state. Other components may intensify envy since they make the whole state more painful. The less significance people attach to their immediate environment and the more significance they attach to their position in overall society, the more likely is envy to appear in what is usually considered to be large subject-object gaps. For these people, more varieties of inferiority are emotionally significant.

---

[26] See my "Why Did Psammenitus Not Pity His Son?" *Analysis*, L (1990): 118–26; and "Envy and Jealousy"; see also A. Tesser, "Self-Esteem Maintenance in Family Dynamics," *Journal of Personality and Social Psychology*, XXXIX (1980)· 77–91.

[27] P. J. Hoffman, L. Festinger, and D. H. Lawrence, "Tendencies Toward Group Comparability in Competitive Bargaining," *Human Relations*, VII (1954): 141–59, p. 144, Dakin and Arrowood.

Although emotional proximity is obviously an important variable of emotional intensity, a precise characterization of this variable is difficult. I now attempt to clarify this issue somewhat. It may be said that greater emotional proximity implies greater emotional intensity (or sensitivity). Such a contention does not provide us with much information. If *S* is emotionally closer to *X* than to *Y*, then for *S* what happens to *X* is emotionally more significant than what happens to *Y*. Hence, *S* is emotionally more sensitive toward *X*; in other words, *S*'s emotional attitude toward *X* is generally speaking more intense than toward *Y*. The discussion would be more informative if we were to discuss at least the following major issues: (a) what constitutes emotional proximity, and (b) the different kinds of correlation between emotional proximity and emotional intensity which are typical of different emotions. Discussion of the first issue may reveal the relationship between proximity in general and emotional proximity, thus clarifying the nature of emotional proximity. Then we shall be in a position to discuss the second issue, which has many practical and theoretical implications.

We may distinguish three major components of emotional proximity: (a) similarity in background, for example, education, age, place, time, and opportunities; (b) closeness in current positions, for example, status, salary, or possession of a certain object; and (c) relevancy for self-evaluation. All components determine the subject's reference group. The first determines the general comparative framework, whereas the borderlines of the specific reference group are also determined by the gap in current positions and relevancy. Those who are similar to us are the natural candidates for social comparison underlying envy. Since there are divergent properties in light of which the comparison can be made, however, the question who is included in the reference group is to a certain extent up to the subject. Thus, we often exclude from our reference group those who are significantly above us though they may be quite similar to us in other aspects. The second factor, namely, the gap in current positions, is in this sense more objective; it expresses, for example, objective differences in salary or rank.

The inclusion in the reference group is then of great emotional significance. Accordingly, we should distinguish between within-group (of reference) inequality and between-groups inequality. Within-group inequality concerns gaps relating to people with whom the subject can compare and conceive of replacing; between-groups inequality refers to gaps related to people who are irrelevant to the subject's self-evaluation and with whom comparison is negligible. The suggested negative correlation between the intensity of envy and the subject-object gap refers to people belonging to two different groups. Concerning within-group inequality, we may envy the person with significantly better fortune more strongly than the one with slightly better fortune. A private in the army would usually envy a corporal or a sergeant and not a field marshal, since a field marshal is not part of his reference group. Within the group of people closely related to him, however, this person may envy a sergeant or even a lieutenant more than a corporal. Similarly, a worker in a big firm may envy the president of the firm less than those working with her; but within the group of people working with her, she may envy those whose yearly income is $1,500 larger than hers more than those earning $100 more than she does. Big gaps in current position or salary may prevent us from considering the object a candidate for our reference group despite great similarity in background. If despite the gap the object is nevertheless in the subject's reference group, the similarity in background should be significant. Within the reference group, members are quite similar to each other and hence are perceived to deserve a similar fortune. In this case the extent of inferiority is the major variable determining the intensity of envy. Outside the reference group there is not much similarity in background, the comparison is harder and the subject's inferiority is not the major factor in determining the intensity of envy.

The inclusion in a given category is of importance in other realms as well. Thus, equal-sized physical differences between stimuli are perceived as smaller or larger depending on whether the stimuli are or are not in the same category. Two different shades of green look more like each other than like any shade of yellow, even when there is no greater difference in wavelength between the yellow and one of the greens than between the two shades of green.[28] In perceiving colors, inclusion in the same reference group (or category) reduces the subjective significance (and hence a given distance within the category is less significant than the same distance between a member of the category and a nonmember); whereas in emotions like envy, such inclusion increases the subjective significance (and hence a given distance within the category is more significant than the same distance between a member and a nonmember). This is compatible with the claim that emotions are involved and interested states concerning those related to us.

---

[28] S Harnad, *Categorical Perception* (New York: Cambridge, 1987). See also my *The Perceptual System: A Philosophical and Psychological Perspective* (New York: Lang, forthcoming).

After clarifying the nature of emotional proximity, we are in the position to discuss the correlation between emotional proximity and emotional intensity. It is obvious that the correlation is different for different emotions. Thus, in love we should expect a positive, linear correlation between emotional proximity and the intensity of love. It seems that emotional proximity is the essence of love. There is, however, a nonlinear correlation between the intensity of love and similarity in background, which constitutes one component of emotional proximity. Empirical evidence thus suggests that love is neither typical of little nor of considerable similarity.[29] In describing the intensity of envy we should take into account the three major components of proximity, namely, similarity in background, closeness (or gap) of current positions, and relevancy. If we take the gap and relevancy as constant, we should expect a more or less positive correlation between similarity and the intensity of envy. A more complex correlation is that between the intensity of envy and the subject-object gap (given that relevancy and similarity are significant and constant). Without drawing the precise lines of this correlation, I may say that the curve depicting it is somewhat similar to a bell-shaped curve. Up to a certain width, envy is quite low. From this point the gap begins to be large enough to generate significant emotional impact. Immediately beyond this point there is a sharp increase in the intensity of envy. There is also a point of maximal envy after which a widening of the gap will reduce envy. The existence of such a point is illustrated by the fact that we are less envious of an increase in the income of rich people than of a corresponding increase in the income of people who are richer than us but poorer than the rich man. In very great gaps an additional increase will not affect the intensity of the subject's envy. In any case, for most people some level of envy still exists on any positive level of the gap.[30]

It is not immediately clear whether envy is more painful in large or small gaps. Some of the factors responsible for the more painful nature of large gaps are: (a) it is harder to ignore or explain away large gaps; (b) the subject's inferiority is more emphasized; hence, depression and despair are more likely to appear; and (c) there is no hope of changing the situation; inferiority is not temporary and the subject is impotent and passive with regard to the possibility of changing the situation. The following factors are more dominant in small gaps: (a) the subject's inferiority is often perceived as a greater injustice since it is considered as stemming from unjust deeds of human beings; (b) the variable of controllability is stronger, that is, the subject's personal responsibility for her failure is more emphasized; hence, frustration, self-blame, and regret are more likely to appear; and (c) an alternative is more available and the feeling of a missed opportunity is stronger.

---

[29] D. Byrne and S K. Murnen, "Maintaining Loving Relationships," In R. J Sternberg and M. L Barnes, eds., *The Psychology of Love* (New Haven: Yale, 1988)

[30] A more precise formulation of the curve is to be found in my *Emotions toward Others*. See also S Paul, "An Index of Relative Deprivation," *Economics Letters*, III (1991): 337–41

The complexity in comparing the intensity of envy in small and large gaps is also represented by the types of available mechanisms for defending the agent's self-esteem. Some mechanisms are more effective in small and some in large gaps. The major mechanisms available in small gaps are: (a) devaluation of the desired item; (b) devaluation of the envied person; and (c) assuming that the subject's inferiority is temporary, either because the subject's genuine value is soon to be revealed or other fundamental factors are about to change the situation. Although to a certain extent these mechanisms are also available in large gaps, they are less effective there since it is harder to devaluate large gaps. A major mechanism, available mainly in large gaps, is that in which we do not choose those obviously above us as friends and acquaintances.[31] This mechanism is less useful in the case of small gaps, typical among friends and acquaintances. Our friends and acquaintances are people in our immediate social and economic environment; precisely these are typical objects of envy. One mechanism that seems to be effective in both cases is where the subject claims that the object's superiority is not worth the sacrifice it requires.

The variety of conflicting factors participating in envy, its painful nature, and the available defense mechanisms make it difficult to determine precisely whether envy is more intense and painful in small or large gaps. The following general trends may be suggested, however: (a) envy is less frequent in large gaps; and (b) if it does arise in such gaps, it tends to be either hardly painful or very painful. Envy is less frequent in large gaps for the following major reasons. First, in large gaps the two people do not belong to the same reference group and hence comparison is harder. Second, there are fewer people whom we consider profoundly superior to us. Third, since it is harder to cope with envy stemming from the profound inferiority typical of large gaps, there is a tendency to transpose profound inferiority into minor inferiority typical of small gaps. Fourth, in many cases where the subject-object gap is large, envy may become admiration.[32] The second claim, namely, that if envy exists in large gaps it tends to be either hardly or very painful, can be explained by taking account of the marginal role that a comparative framework has in such cases. Since comparative considerations are hardly present, envy is similar to wishful thinking and is not painful. If we insist on taking comparative considerations into account in large gaps, then our situation is very painful since it involves profound inferiority. In *Amadeus*, Salieri's envy toward Mozart is an example of how painful and destructive envy involving a large gap can be. Indeed, envy of natural gifts often turns into hatred since in both cases the object's basic and permanent characteristics are negatively evaluated. The issue of desert, which is more pronounced in great proximity, where there is no apparent reason for the subject's inferior situation, can also play a role in large gaps. In this case the complaint will be more general, often directed at impersonal circumstances. Salieri does not doubt Mozart's significant superior standing in music, but he doubts God's judgment in bestowing a great musical talent on such a rude and ugly person as Mozart.[33] Hostile or malicious envy, which is the prototypical kind in pathological cases discussed in clinical literature on envy, is more common in large gaps. We should not reduce envy to hostility, however, or vice versa. Envious people dislike their inferior position; they do not necessarily feel hostility or dislike toward the envied person.

---

[31] See also Elster, "Envy in Social Life," manuscript (1991); Montaldi.

V

I now turn to discuss the sociological aspect of the correlation between envy and inequality. Although inequality refers primarily to gaps in current positions, the other factors of proximity, namely, similarity in background and relevancy, also play some role in determining the level of inequality. I shall examine the level of envy in conditions of reduced social and economic inequality, rather than perfect equality, since the latter is an ideal of doubtful feasibility (and, some may add, moral grounds); anyway, it is not a goal for the foreseeable future. Reducing inequalities is a more feasible goal which to some extent has already been realized in several societies.

---

[32] Envy may turn into admiration if the object is excluded from our specific reference group, namely, when the object's success is not relevant in determining our worth. A typical case of admiration involves high relevancy of the specific domain and little similarity to the specific person: we usually admire those in our domain who are not similar or close to us and hence do not threaten our self-esteem. In these cases the object's superiority is typically perceived to be unalterable. Another attitude in this respect is when we perceive the object's good fortune to increase our esteem. This attitude, which typically involves low relevancy of the specific domain and little similarity to the specific person, has been characterized as basking in reflected glory; it is common, for example, in sport People feel that they can share in the glory of a successful other with whom they are in some way associated By being associated with someone's success, the subject's social as well as self-image are thought to be enhanced. See R. B. Cialdini, R J. Borden, A. Thorne, and L R. Sloan, "Basking in Reflected Glory. Three (Football) Field Studies," *Journal of Personality and Social Psychology*, XXXIV (1976). 366–75.

[33] See also R. Smith On the relationship between envy and hate, see my "Anger and Hate," *Journal of Social Philosophy*, XXIII (1992).

The central concern in envy is different from the egalitarian moral concern that calls for the reduction or even elimination of different inequalities. No doubt, by fulfilling the desire underlying envy, namely, gaining what someone else has, I may become equal to this person in this respect, but this is not what the egalitarian moral concern amounts to. As suggested above, envy is not a moral emotion. Envy differs from the egalitarian moral concern in at least two major ways. First, it involves a partial rather than a general concern: the envious person is not concerned with equality as a general value; the claim to equality is merely a desire to improve the subject's personal situation and thus does not appear when inequality favors the subject. Second, envy also surfaces in cases where the demand for equality is unrealizable and has nothing to do with egalitarian moral principles; for example, when one envies the beauty or wit of another person. The claim that *egalitarianism is a central concern in envy* should be rejected. Although this claim is different from the claim that *envy is a central concern in egalitarianism* or that *egalitarianism is motivated by envy*, these claims should also be rejected. I do not discuss these views here, but merely repeat that egalitarianism is a general, moral stand in a way that envy is not.[34]

A more complex issue concerns the suggested negative correlation between reduced inequality and the intensity of envy. This correlation holds for most types of inequality; a notable exception is equality of need satisfaction. Unlike most other kinds of equality, this kind does not express a quantitative identity but a qualitative similarity that is not associated with a continuous comparison with others. This kind of equality increases our overall satisfaction and is therefore more likely to reduce envy. The problem is that this is the hardest kind to achieve and it is difficult to determine its presence, especially under conditions of scarcity. Concerning other types of equality, only perfect equality can significantly reduce envy. (Equality in need satisfaction is the closest to perfect equality.)

It is not easy to provide empirical confirmation to the suggested correlation between envy and inequality. The notion of "emotional intensity" is not unitary; it applies to different phenomena, not all of which are correlated. Emotional intensity can be analyzed in light of the different emotional components, for example, evaluation, cognition, motivation, and feeling, and in each component intensity is expressed in a different manner. Such manners may be divided into two major aspects: (a) magnitude (or peak); and (b) duration.[35] Despite the complexities in defining and measuring emotional intensity, there are empirical scales for measuring some emotions such as love, anger, and jealousy. Measuring the intensity of envy is especially complex, since a person's envy is specific to people and domains of personal importance. It is also difficult to measure the intensity of envy in a certain society, as there are no general behavioral rules that express the presence and intensity of envy unequivocally. Nevertheless, there is some empirical evidence suggesting that the inclination toward equalization of fortunes is strongest when the subject and the object are more or less on equal footing, and that envy does not diminish with reduced inequality but often even increases. One type of such evidence refers to the degree of satisfaction of Black Americans from different socioeconomic levels. It was found that as Blacks move up the economic and social ladder, they become less satisfied and more likely to take part in protest or riot activities. There is further evidence supporting the relative deprivation theory which claims that "stratified societies can tolerate higher inequality than unstratified societies. As people become more engaged with each other, they have less tolerance for a given level of inequality."[36] In what follows, I offer a further step in clarifying the complexity of the envy-inequality correlation. I do not present results of empirical investigations, but analyze the level of envy in one of the most egalitarian contemporary societies, the Israeli kibbutz. This analysis illustrates the role in the generation of envy of the factors discussed above.

---

[34] See Schoeck, chs. 14–5, for references to such views; see R Young, "Egalitarianism and Envy," *Philosophical Studies*, LII (1987) 261–76, for a rejection of them. See also G. F. de la Mora, *Egalitarian Envy* (New York. Paragon, 1987).

The Israeli kibbutz is by and large an egalitarian society whose members spend a great deal of their life together and enjoy equal standing with respect to the basic needs of food, health, education, and access to culture. Inequality is minimal and is usually represented by different quantities of luxury items or in the type of work done by each member. It is instructive to find out whether the significant reduction of inequality in this society has led to a reduction in the level of envy. I first indicate characteristics of kibbutz society which increase the level of envy.

---

[35] See N. H. Frijda, A. Ortony, J Sonnemans, and G. Clore, "The Complexity of Intensity Issues concerning the Structure of Emotion Intensity," in M. S. Clark, ed , *Emotion: Review of Personality and Social Psychology*, XIII (1992): 60–89; my *Emotions toward Others*.

[36] S. Yitzhaki, and R. I. Lerman, "Income Stratification and Income Inequality" (working paper, Department of Economics, The Hebrew University of Jerusalem, 1989). See also Crosby; Heider, pp. 288–9; Schoeck.

1. There is great similarity in the members' background and current situations and there are small differences in the members' monetary resources and ways of life. When easily perceived differences are infrequent, smaller differences become easier to perceive and acquire greater emotional significance. Moreover, when a society is not large or flexible enough to permit segmentation into groups, it might not be able to maintain a variety of opinions, abilities, and fortunes, thereby increasing similarity and hence envy.
2. Personal encounters are frequent, making it difficult to forget prevailing envy: the dynamic of "out of sight, out of mind" does not exist. (The size of each kibbutz varies from a population of about 100 to about 1,500, small enough for frequent encounters and keeping people informed about the fortunes of each one.)
3. Since the background and opportunities of each member are more or less similar, one can only blame oneself for one's inferior position. Greater personal responsibility usually intensifies emotions.
4. Envy has more legitimacy than in other places. There are two major reasons for this. (a) Egalitarian ideology justifies condemnation of inequality and thus may be perceived as implying some justification of envy. Moreover, because of the predominance of the egalitarian ideal, people are often preoccupied with equality for its own sake, thereby decreasing their readiness to be satisfied with their own situation and thus facilitating the generation of envy.[37] (b) In small subject-object differences, the subject feels more entitled to have what the object has than in large differences.
5. There is a stronger and more justified feeling that the other's gain means the subject's loss. This feeling, which is not a necessary feature of envy (it is more so of jealousy), is common in the kibbutz where the gross income is shared by all members; if one gets more than other members, these members are in a certain sense supporting this person.
6. Envy can have practical results: the society is actually capable of demoting envied people from their position. This can be done either by making them less welcome from a social viewpoint, or by not approving many of their requests.

The first three characteristics express the great proximity in the kibbutz. The next two characteristics suggest why the element of desert is stronger in this society. The last characteristic indicates that envy has practical results. All characteristics increase the availability of an alternative, namely, the likelihood of reversing the subject's present inferior situation; hence, the intensity of envy increases.

---

[37] See also H. Frankfurt, "Equality as a Moral Ideal," in *The Importance of What We Care About* (New York: Cambridge, 1988), p. 135.

Some characteristics of kibbutz society may reduce envy, however. The major ones are the following.

1. Kibbutz society is less competitive in many areas. Thus, the kibbutz's egalitarian ideology attaches less importance to economic and professional standing
2. An envious person is usually ignorant of many of the object's qualities, focusing attention only on what is envied. The desired qualities receive therefore disproportionate attention from the envious person. In the kibbutz, this distortion is less likely to occur because there is more intimate acquaintance with all qualities, including the shortcomings, of the envied person. As suggested above, a broader perspective usually decreases emotional intensity. (Of course, the more intimate acquaintance may induce further envy as when the object is superior to the subject in these areas as well. Indeed, in classical antiquity it was assumed that whoever excels in one respect excels in all. But since no one is perfect, it is always possible to find shortcomings and to assign them disproportional weight.)
3. An individual's social circle is not limited to those of a similar economic and professional standing—the primary candidates for our envy.
4. Reducing inequalities often leads to an increase in self-satisfaction —at least for those in inferior positions, who are the ones who experience envy. Since the absence of self-satisfaction is an important component in envy, its increase may lead to a lower level of envy.

The above analysis indicates that even in this small and largely homogeneous society, different factors have conflicting influences on the level of envy. These conflicting factors make it hard to determine the comparative level of envy in each kibbutz with certainty. Moreover, personal and contextual factors typical of each particular kibbutz would often have to be taken into account in order to determine the exact level. Although the main purpose of the above analysis was to illustrate the presence and nature of the factors underlying envy, it can also serve to indicate the comparative level of envy in this egalitarian society. The major factors raising the intensity of envy in kibbutz society are greater proximity (concerning both similarity in background and closeness in current positions), stronger desert claims, and greater availability of an alternative. The major factors reducing envy are less competition, broader personal acquaintance and less profound inferiority. In light of our analysis of envy and emotional intensity, it seems that the first group has greater weight; hence, envy is likely to be stronger in this society.[38]

A similar analysis can be conducted with regard to modern Western society in general. I do not intend to carry out such an analysis here but to point out that the level of envy seems to increase also in this society. The major factors responsible for this are: (a) modern Western society is less stable than traditional societies, and instability is a major variable in the rise of emotions;[39] (b) modern Western society is much more competitive than its predecessors; (c) the communications revolution brings people closer to each other; and (d) ideological condemnation of huge social and economic gaps is increasing—the present hierarchy is no longer accepted as a justified, natural phenomenon. There are of course conflicting factors, for example, improved living conditions and greater stress on personal development. I cannot compare the weight of these and other factors here, but again the factors responsible for increased envy seem to be more central.

The presence of a mainly negative correlation between inequalities and envy does not imply that we should not strive to reduce inequalities; reducing inequalities may have positive results of a different sort. We should be aware of the consequences (vis-à-vis envy) of such a development, however. Reducing inequalities may lead (as Rawls argues) to a more just society; but (contrary to Rawls) this society may witness a rising level of envy. When social and economic gaps are large, the chances for rage, hate, frustration, and several types of violent reactions are greater. When these gaps narrow, the intensification of envy usually follows, along with a reduction in these attitudes. If we are to witness a further reduction in social and economic inequalities, we should expect the problem of envy to become more prominent. Whatever social and moral advantages reduced inequality has, reduced envy is not one of them.

VI

What are the possible ways of reducing envy? One obvious way is to increase the subject-object distance, namely, to decrease the number of people who are of emotional significance to us. It seems, however, that reducing envy is not worth the price of widening the subject-object disparity in this manner. The harmful consequences would exceed the benefits. Another related way is to reduce the availability of an alternative. Again, reducing envy is not worth this price of limiting one's opportunities which may reinforce despair and indifference and debilitate hope in human life.

---

[38] Kibbutz ideology is aimed at equality of need satisfaction; the ideal is supposed to be "from each according to his ability, to each according to his need." As mentioned, this equality is the most likely to reduce envy but the hardest to achieve. The areas in kibbutz' life which are close to fulfilling this ideal kind of equality are food, health, and to a lesser extent education. It seems that in these areas, where inequality of need satisfaction is significantly reduced, the level of envy is reduced as well.

[39] In the political realm, changing one's position is facilitated by the democratic structure of modern societies. Isocrates argues that monarchy is preferable to democracy since it involves less envy· the king is so far above his subjects as not to excite envy. See Walcot, p 30

Envy can be reduced by making social comparison, and hence competitiveness, less significant. Two major means for doing it are diversifying valued domains and settling for less. The first means is that of expanding the range of valued domains in which people can perform well compared to others.[40] Thus, chances increase that more people will enjoy superiority in at least one domain. Since we can expect people to consider this domain to be quite relevant in assessing their value, we can expect the reduction of envy and competitiveness. Diversifying valued domains requires, then, two elements: (a) adding new types of domains a person can be satisfied with; and (b) the realization that each domain has its own intrinsic value. Achieving these objectives, which involves stressing the intrinsic value of people (including oneself), increases dissimilarity among individuals; hence, it offsets the increased similarity stemming from reduced socioeconomic gaps.

The second means to lessen the significance of social comparison is that of settling for less. In this case another person's success will bother us less. As A. Parducci[41] remarks, "With lower standards, regular accomplishments would be much more satisfying." Take the case of a man who has an exalted picture of the ideal woman in his imagination, a standard that no real-life woman could meet. Because of such exalted standards, all his romantic relationships are doomed to be disappointing. If this man lowers his standards, he can be happier. What now appears to him as "the best" may be a real-life woman, with all the faults that previously seemed too grievous to bear. If we can truly settle for less, happiness is likely to increase and envy is likely to decrease since we are more likely to gain what we want. Some societies, such as the kibbutz, foster the ideal of being content with little, which is similar to the ideal of settling for less.

---

[40] See R. Nozick, *Anarchy, State, and Utopia* (New York: Basic, 1974), R. H Smith, E. Diener, and R. Garonzik, "The Roles of Outcome Satisfaction and Comparison Alternatives in Envy," *British Journal of Social Psychology*, XXIX (1990): 247–55

[41] "Value Judgments: Toward a Relational Theory of Happiness," in J. R Eiser, ed., *Attitudinal Judgment* (New York: Springer, 1984), pp 14–5 Settling for less is not recommendable when a higher level of success can really be achieved; in this case, Parducci argues, a conflict between achievement and happiness arises.

Both attitudes realize a similar evaluation concerning the limited value of the continuous race to win. But whereas the first ideal addresses absolute deprivation, the second concerns relative deprivation. Hence, for the purpose of reducing envy, the second ideal is more appropriate.

We may combine the suggested two means to reduce envy by speaking about settling for being different. In some cases different may be less, but in others it may just refer to a noncomparable domain. Adopting these two means represents a kind of maturity of the subject (and society) in which social comparisons are less significant in determining the subject's happiness. Such determination becomes more intrinsic to the subject's own activity. Other means to reduce envy are those effective in reducing emotional intensity in general; for example, making one's perspective less partial. Although these means are not easy to employ, they have a good chance of reducing envy, especially if we realize their role in increasing our happiness.

In summary, I have argued that envy is not a moral emotion and that the subject's inferiority is a central concern in envy. This is not the only concern, however, since there is no positive correlation between envy and inferiority. Another important concern is that of desert. Accordingly, the focus of concern in envy is the subject's undeserved inferiority. This concern does not imply a positive correlation between envy and the subject-object inequality, since questions of desert are more prominent regarding those who are similar and close to us, namely, comparable with ourselves. Determining the type of this correlation should take into account whether the object is within or outside the subject's reference group. Envy is more frequent within the reference group where the correlation between the intensity of envy and subject-object inequality is usually positive. Outside the reference group, the correlation is often negative since greater inequality makes the subject-object comparison less significant. This correlation holds also on the sociological level. Accordingly, we can expect that reduced socioeconomic gaps will increase envy. Major means for reducing envy are those making social comparison less significant.

AARON BEN-ZE'EV

University of Haifa

THE IDENTITY THEORY OF QUOTATION*

Suppose I want to talk about a bit of language, say, the title of this essay. How might I go about doing this? I can give it a name. '*U*' will do. I can then say that *U* has twenty-eight letters, and what I say will be true. Another option available to me is to offer a description. I can describe the title in terms of its constituent words. I can give the names '*V*', '*W*', '*X*', '*Y*', and '*Z*' to the five words that compose it. I can then say the expression consisting of *V* followed by *W* followed by *X* followed by *Y* followed by *Z* has twenty-eight letters, and again what I say will be true. Yet another option is to demonstrate it. I can say, pointing to the top of the page, that *that* expression has twenty-eight letters, and once again I shall say something true.

Naming, describing, and demonstrating are distinct ways of mentioning the title. Together they exhaust the standard ways of mentioning nonlinguistic objects. But most of us would balk at the idea that our options for mentioning language are limited to these three. Surely, if I want to mention the title, I can simply quote it. I can talk about the title by saying that 'The identity theory of quotation' has twenty-eight letters, and what I say will be true. But in quoting the title, am I doing something other than naming it, describing it, or demonstrating it? This is the main question which I want to address. The answer I suggest is "yes." I argue that the use of expressions in quotation is fundamentally different from nonquotational uses of language.

I

In natural language, quotation is a way of mentioning things. Its use is easy to learn and once learned it is readily applied and understood. These qualifications make quotation a phenomenon worthy of explanation. The fact that quotation, like the modalities and the attitudes, gives rise to opacity makes it a problem for the philosophy of language. As such problems go, however, quotation would seem to be on the easy side. Its workings appear to be a paradigm of simplicity: in quotation, it seems that words are used and mentioned in the same breath. Moreover, this correlation would appear to be explanatory: it seems that words in quotation are mentioned simply by being used.

---

* I have received helpful comments from John Perry, Jamie Rucker, Michael O'Rourke, Henry Smith, Johanna Carr, and a referee for *The Philosophical Review*.

0022-362X/92/8911/582–605

In what follows, I shall argue that, where quotation is at issue, the appearances are the reality. I shall argue that, in quotation, expressions are used to mention themselves and that this correlation is rule-governed. I call this view the *identity theory,* which is based on the idea that linguistic expressions may be used in different ways. According to the theory, a special *quotational use* of expressions underlies the phenomenon of quotation. The quotational use of an expression contrasts with what I call its *standard use.* The paradigm of standard use is the use of expressions outside of quotation in unembedded contexts. The contrast is illustrated in (1):

1. 'Three words long' is three words long.

The first use of the expression 'three words long' is quotational, while the second is standard. The first is interpreted by the principle of identity, the second is not.

Alfred Tarski, W. V. Quine, Peter Geach, and Donald Davidson are among those who find the appearances misleading. These writers maintain that while quoted expressions may seem to mention themselves, the identity is illusory and, in point of fact, the expressions used are not what is mentioned. They also deny that the principle of identity or any related principle plays a role in the interpretation of quotation. They recognize only the genera of uses that I call standard, and advocate theories that treat quotation as involving a standard, though perhaps disguised, use of names, descriptions, or demonstratives. The quotation in (1) is thought to be analogous to (2), (3), or (4):

2. *A* is three words long.
3. The expression consisting of *B* followed by *C* followed by *D* is three words long.
4. *That* expression is three words long.

(Where '*A*', '*B*', '*C*', and '*D*' are names and 'that' is a demonstrative.)

I call these theories "standard." A *standard theory* is one that analyzes quotation on the model of a standard use of an expression. Below I shall explain how standard theories differ from the identity theory and attempt to show why the commitment to standard semantic concepts makes it impossible to give a correct answer to what would seem to be a very simple problem. My choice of the term 'standard', which I contrast with 'quotation', is meant to underscore the essential difference between quotation and other uses of language.

It will help to prevent confusion later on if I take a moment to formulate some terminology and draw some distinctions. At the outset, it is important to distinguish expressions from their tokens. The former are types, which are abstract particulars. The latter are tokens, which are concrete particulars. When I say 'expression' I mean expression type, and when I say 'token' I mean expression token. I call an expression a "mentioning expression" if it is *used* to mention something. Being a mentioning expression is a property that an expression has on a particular use. My use of the verb 'mentioning' is meant to be neutral between different ways of talking about objects. I use 'semantic value' as a term for the contribution an expression makes to the interpretation of sentences of which it is a part. I use 'object mentioned' or 'mentioned object' as terms for objects that are mentioned. Objects mentioned are the semantic values of mentioning expressions. When I say that the semantic value of an expression has been 'stipulated', I mean that the expression acquired its value as a direct result of the performance of a particular act or acts. I do not mean to imply that the acts were performed with the intention of giving the expression a value.

To give a precise analysis of the semantics of quotation, three basic questions have to be answered:

5. What expressions are used and what roles do they play?
6. What type of object is mentioned?
7. How is the semantic value of the mentioning expression determined?

I shall let these questions guide my discussion. In the next section, I review some well-known standard theories of quotation and the answers they give to questions (5)-(7). In section III, I briefly summarize the identity theory and the answers it gives to questions (5)-(7). Sections IV-VI address questions (5)-(7) in detail. I provide reasons why they should not be answered in the manner of the standard theories and why they should be answered in the manner of the identity theory, taking the occasion to go more deeply into details of the proposals.

II

Question (5) asks us to determine the parts of a quotation, if there are any, and assign credit for overall effects where credit is due. Given the orthographic structure of standard *written* quotation, the candidates for the mentioning expression are limited to three. The mentioning expression could be (a) the expression consisting of the marks and what stands between them, which I term the 'quotation as a whole'; (b) the expression between the marks, the 'quoted expression'; or (c) the quotation marks alone. Tarski, Quine, and Geach have advocated the view that it is the quotation as a whole which mentions an object, while Davidson is responsible for the novel suggestion that the quotation marks are complete mentioning expressions in themselves and that the quoted expression is not used at all.

A view held by Tarski, Quine, Geach, and many others regarding what is mentioned in quotation, question (6), is that quotation is a device for talking about linguistically interpreted types. These types come in different varieties. On one theory advocated by Tarski and Quine, these types are concatenations of letters. Geach's explicit position is that the object mentioned is a concatenation of words. Taking a subtly different stance, Davidson holds the objects mentioned by quotations to be shapes.[1] Davidson and the advocates of the theory of interpreted types are united in maintaining that in quotation we do not talk about tokens. Disagreements between Davidson and type theorists concern the boundaries of the types and whether the types are linguistically interpreted; a linguistically interpreted type may encompass many distinct uninterpreted shapes.

Question (7) asks about how the value of a quotation is fixed. The answer must give an account of how it is determined which object a given use of quotation mentions. In this connection, Tarski and Quine have promoted what has come to be called the "name theory." This is the view that quotations are standard occurrences of semantically simple expressions.[2] As Tarski remarks:

> Quotation-mark names may be treated like single words of a language, and thus like syntactically simple expressions. The single constituents of these names—quotation marks and the expressions standing between them—fulfil [sic] the same function as the letters and complexes of successive letters in single words. Hence they can possess no independent meaning. Every quotation-mark name is then a constant individual name of a definite expression (the expression enclosed by quotation marks) and in fact a name of the same nature as the proper name of a man (*op. cit.*, p. 159).

An expression is semantically simple if its standard semantic value is not fixed by general principles. Words like 'and', 'car', and 'hello' are semantically simple. To say that a quotation is a standard occurrence of a semantically simple expression is to say that its value is not given by general principles, that some stipulation must have been made for it to have obtained its value.

---

[1] Davidson seems to be influenced by Tarski, who held that quotations refer to classes of expressions of like shape. See Tarski, "The Concept of Truth in a Formalized Language," in *Logic, Semantics and Metamathematics* (New York: Oxford, 1969), p. 156n.

[2] Jonathan Bennett reports that Quine has maintained, in personal communication, that in his writings he uses the term 'name' as a catchall phrase for expressions that are semantically simple—"Quotation," *Noûs*, XXII (1988): 399–418

Along other lines, Tarski (*op. cit.*, p. 160) and Quine[3] have also suggested that quotations be analyzed as standard uses of collapsed descriptions, which describe expressions by their spelling. This theory, which has been called the "spelling theory," treats the quotation 'hello' as a short form of the description "the expression got by writing '*h*' followed by '*e*' followed by '*l*' followed by '*l*' followed by '*o*'." Another version of the description theory, given by Geach,[4] is identical except that the basic parts of descriptions mention individual words rather than letters. He writes: "I maintain that the quotation 'man is mortal' is rightly understood only if we read it as meaning the same thing as " 'man' & 'is' & 'mortal'," i.e. read it as *describing* the quoted expression in terms of the expressions it contains and their order" (*ibid.*, pp. 82-3). I call this the "lexical" description theory. It is important to note that neither the spelling nor the lexical theory gives an account of the semantics of the basic parts of the descriptions. These theories treat the parts as semantically simple, a subtlety which is obscured by the use of quoted letters or quoted words.

The range of views is represented in the following chart:

Proponent	semantic category	mentioning expression	method of fixing	mentioned object
T/Q	name[5]	quotation as a whole	stipulation	expression
T/Q	description	quotation as a whole	stipulation for parts	series of letters
Geach	description	quotation as a whole	stipulation for parts	series of words or unambiguous expression
Davidson	demonstrative	quotation marks	demonstration	shape
Frege	anything	quoted expression	principle of identity	expression
Searle	anything	quoted expression	principle of identity	expression
CGW	anything	quoted expression	principle of identity	expression

CHART A

---

[3] *Word and Object* (Cambridge: MIT, 1964), p. 212.

[4] *Mental Acts* (New York: Routledge, 1957).

[5] See fn. 2.

III

With regard to the issues raised by question (5), the identity theory diverges from Tarski, Quine, Geach, and Davidson, while agreeing with John Searle and Gottlob Frege. In "On Sense and Meaning," Frege[6] writes:

> If words are used in the ordinary way, what one intends to speak of is what they mean. It can also happen, however, that one wishes to talk about the words themselves or their sense. This happens, for instance, when the words of another are quoted. One's own words then first designate words of the other speaker, and only the latter have their usual meaning. We then have signs of signs. In writing, the words are in this case enclosed in quotation marks. Accordingly, a word standing between quotation marks must not be taken as having its ordinary meaning (*ibid.*, p. 201).

In this passage, Frege is primarily concerned with direct and indirect discourse reports. His comments about direct discourse also apply, however, to uses of quotation where words are mentioned and discussed but not attributed to a person as something said. Frege's view is that the use of an expression on the interior of a set of quotation marks is special. Quotation marks, which are most common in written speech, signal this new use. John Searle[7] puts the point succinctly when he writes that in quotation "a word is uttered . . . but not in its normal use. The word itself is *presented* and then talked about, and that it is to be taken as presented rather than used conventionally to refer is indicated by the quotes" (*ibid.*, p. 76; my emphasis). The quotation as a whole is analyzed into the marks that signify quotational use of the quoted expression and the quoted expression that is itself used to mention an object. All expressions, even those whose standard uses are not as mentioning expressions, become mentioning expressions in quotation.

As regards the question of what quotation mentions, raised in (6), I maintain that the most common use of quotation is to mention linguistic types. These types may be interpreted or uninterpreted, like 'man' or 'ggrmph'; they may be structured or unstructured, like 'this man' or 'man'. If uninterpreted, however, they must at least be potential bearers of linguistic characteristics. In reply to question (7), I hold that a quoted expression is related to its value by identity: a quoted expression mentions itself. The identity theory's distinction between standard and quotation uses can be represented schematically as follows:

---

[6] In *The Philosophy of Language*, Martinich,ed. (New York. Oxford, 1985)

[7] *Speech Acts* (New York Cambridge, 1989).

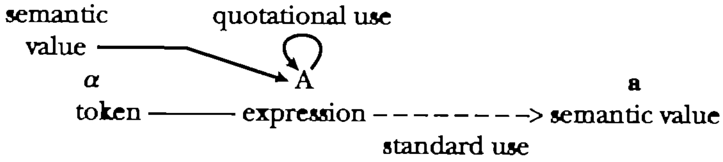


CHART B

A solid line marks the fact that the relation is determined by general principles. I use a dotted line to connect the expression with its standard semantic value to indicate that the connection is determined by stipulation. Stipulations are not general. Since a stipulation may or may not have been carried out, a basic expression may or may not have a standard semantic value. What is commonly called the "lexicon" of a language consists of those basic expressions which have standard semantic values.

IV

What are the parts of a quotation and what role do they play? An important fact about quotation which seems to have been overlooked by those who say that the quotation marks either are the mentioning expression or form an essential part of the mentioning expression is that quotation marks may be absent. In conversation, oral promptings ("quote-unquote") or finger-dance quotes can often be omitted without impairing the intelligibility or well-formedness of the utterance. When I introduce myself, I do not say "my name is quote-unquote Corey," nor do I make little finger gestures or even use different intonation in order to show that it is my name and not myself that is being talked about. Linguists do not say "quote-unquote flying planes can be dangerous is ambiguous." Philosophers do not say "the belief report quote-unquote . . . ." In general, when it is clear that the subject matter is linguistic, we simply utter the expression without marking it in any special way.

These mundane facts show that Tarski's, Quine's, and Geach's proposals do not apply to the vast majority of spoken uses of quotation. These writers claim that the quoted expression is a mere accident in the course of the spelling of a longer word, which contains in addition the two quotation marks.[8] It is this longer word, the quotation as a whole, which is the mentioning expression. But when quotation marks are lacking, the quotation as a whole does not occur.

---

[8] Quine, "Reference and Modality," in *From a Logical Point of View* (New York: Harper & Row, 1961), p. 140.

So, on this proposal, most quotations do not exist—a conclusion that is difficult to accept.

Davidson's[9] thesis is vulnerable to the same counterarguments. He conceives of quotation marks as referential expressions, roughly synonymous with the locution, 'the shape of *that*':

> On my theory which we may call the demonstrative theory of quotation, the inscription inside does not refer to anything at all, nor is it part of any expression that does. Rather it is the quotation marks that do all the referring, and they help to refer to a shape by pointing out something that has it (*ibid.*, p 90).

On Davidson's account, the subject of the sentence in (8) is the quotation marks:

8. 'Hello' is a salutation.

That is, (8) appears as:

8′. ' ' is a salutation.

And when uttered without quotation gestures as:

8″. is a salutation.[10]

Davidson's theory does not provide for a subject to replace the quotation marks when quotation gestures do not accompany the utterance.[11] Unless some other element can be found to serve as the subject of the sentence, the theory makes the manifestly false prediction that spoken uses of (8) are ill-formed.

It has been suggested in defense of Davidson's proposal that the quotation marks in an utterance of (8) are implicit.[12] These implicit quotation marks would then serve as the subject in (8″), saving the structure from ill-formedness. There are circumstances in which a sentence may be perfectly well-formed and yet lack an explicit expression in a subject position. In general, imperatives may go subjectless, as may declarative sentences in languages such as Spanish and Italian ("pro-drop languages"). Contemporary linguistic theories commonly postulate implicit expressions in place of their missing explicit counterparts to explain the syntactic and semantic properties of these constructions. Nevertheless, these theories are careful to place tight restrictions on the contexts in which implicit elements may occur. These restrictions exclude the situation in (8″). The predicate in that sentence cannot license an empty element. Current theory provides no support for the claim that (8″) contains an implicit subject.[13] In the absence of further evidence for the existence of an implicit subject in (8″), Davidson's suggestion that quotation marks are mentioning expressions must be rejected.

---

[9] "Quotation," in *Truth and Interpretation* (New York: Oxford, 1984).

[10] A similar argument against Davidson's demonstrative theory of indirect discourse is given by Gabriel Segal in "A Preference for Sense and Reference," this JOURNAL, LXXXVIII, 2 (February 1991) 73–89.

[11] Indeed, it is difficult to see how gestures, even when present, could serve in a syntactic capacity in spoken or written language.

[12] This is a response which I have often heard. I cannot recall all of the people who have suggested it. Stephen Neale is one

The facts about written quotation are similar. In writing, we can also dispense with quotation marks. We can indicate that an expression is being quoted by writing it on a separate line or in italics. One must be pressed very hard to adopt a Quinean point of view and see the italicized form of the quoted expression or the quoted-expression-on-another-line as distinct from the quoted expression, as if italicization or setting an expression off created a new expression. There is even less reason to attempt to persevere in Davidson's way of looking at things and reify the italicization or the being-on-another-line of the quoted token into a mentioning expression.

Yet writing does differ from speech in some important respects. Unlike spoken quotation, it appears that some signal is required to indicate that the expression is to be quoted and not used as usual. Though standard in speech, (9) is simply not acceptable written language.[14]

9. Hello is a salutation.

The fact that (9) is ill-formed should not be taken as evidence that quotation marks form an essential syntactic or morphological part of the sentence, and hence as evidence for a position like Quine's or Davidson's. Causes of ill-formedness are varied. Written quotations without quotation marks are simply incorrectly punctuated and are deviant in the same way that other incorrectly punctuated expressions are deviant:

10. Did you eat your peas.
11. John who went to school liked to stay up late.

There are a number of ways of explaining how punctuation can effect syntactic well-formedness without supposing punctuation marks to be words or parts thereof. A view I find attractive is that in common written language punctuation marks play a role analogous to the role phrase-structure markers play in linguistics. Just as phrase-structure markers signal that a linguistic element has a certain structure or belongs to a certain category, so too do punctuation marks signal that an expression has a certain linguistic character. Neither are grammatical constituents of the expressions they accompany. If a sentence is followed by an exclamation point, then it is a command or an exclamation; a comma might be a sign of a break between a series of constituents; an expression flanked by quotation marks is a quotation.[15] The absence of the appropriate marks would signal that the expression lacks the characteristics in question. (This is necessary to ensure that expressions without punctuation are not systematically ambiguous between punctuated and unpunctuated readings.) The sentence in (9) would be ill-formed because the word 'hello' would have its normal use as an interjection in a context where a noun phrase is required.

---

[13] It should be clear that any thought of salvaging Quine's theory by supposing empty elements (empty letters, in this case) in place of the missing quotation marks should be met with at least as much skepticism as it is in the case of Davidson's theory.

[14] Sometimes something like this is written:

He is called John.

But here the verb 'called' signals that a quotation is to follow.

I believe the simplest and most natural way to explain these facts is to take the quoted expression to be the mentioning expression. Quotation marks, on this account, are neither mentioning expressions nor parts of mentioning expressions. Quotation marks are punctuation. As punctuation, they signal that the quoted expression has a special use.[16] They take their place alongside the period, semicolon, question mark, and comma.[17] The differences between written and spoken quotation can be attributed to the more formal character of written expression.

Although this account strikes me as the most plausible, it is not the only one that is compatible with the facts. We have seen that there are cases in which quotation marks are present and cases in which they are not; spoken quotation can occur with or without quotation marks, though marks are usually not present; direct quotation in writing may occur with or without marks, though the use must be indicated in some fashion. We could say that there are two sets of mentioning expressions which can be used in quotation, one consisting of expressions with quotation marks (the quotations as a whole) and another consisting of expressions without quotation marks (the quoted expressions). There will be: 'hello' and ' 'hello' ', 'goodbye' and ' 'goodbye' ', 'stop' and ' 'stop' ', and so on. Counterparts in each set will always mention the same object. Although this proposal accords with the facts, it is still undesirable insofar as it leaves unexplained the fact that expressions with quotation marks always mention the same thing as expressions without quotation marks. Another possible explanation is that the quoted expression mentions the object and quotation marks are functional expressions. This would explain the symmetry between expressions with and without quotation marks. The idea of functionless functional expressions goes down hard, however. The fact that these purported functions can be left off when context requires indicates, I think, they are more like context markers (punctuation) than like functions. Given that neither of these proposals is satisfying, I shall stick to the view that the quoted expression mentions the object, while the quotation marks are punctuation.

---

[15] For a detailed discussion of the syntax and semantics of punctuation, see Jeffrey Nunberg, *The Linguistics of Punctuation* (Stanford: CSLI, 1990).

[16] Quotation is sometimes described as a *linguistic context* in which words take on a different significance This way of speaking is not entirely accurate, since contexts are typically marked grammatically, but the grammatical signs of quotation (the marks) are often absent.

[17] Sanskrit exhibits a quotative marker (*iti*) which is placed after the quoted material; see J S. Speijer, *Sanskrit Syntax* (Leiden: Brill, 1886), pp. 379–88:

> nrpo 'ham iti vadati
> King-NOM I-NOM QUOTATIVE-MARKER he-says
> He says "I am the king"
> (I am endebted to Henry Smith for this point.)

According to the *Encyclopedia Britannica*, quotation marks were introduced in the West along with the exclamation point around 1550. Prior to this time, their work was done by contextual factors.

Before I leave this topic, I want to examine an argument suggested by Davidson's article. The argument is designed to show that quoted expressions cannot be parts of quoting sentences. If the argument is sound, then we must give up our contention that the quoted expression is the mentioning expression.[18] The basis of the argument is Davidson's conception of the finiteness constraints on learnable languages. "In an adequate theory," Davidson writes, "every sentence is construed as owing its truth or falsity to how it is built from a finite stock of parts by repeated application of a finite number of modes of combination" (*op. cit.*, p. 83). In other words, for the interpretation of a language to be finite, only a finite number of elements may occur in its sentences and the semantics must be determined by a finite number of finitely statable principles. This condition reflects assumptions about learnability. The number of basic expressions is limited because a speaker could not learn the semantic values of an infinite number of expressions. Sentences of a learnable language cannot contain an infinite number of expressions whose semantic values have been stipulated. Admitting the quoted expression into the quoting sentence, so the argument goes, would lead to a violation of this principle.

---

[18] This argument is closely connected with the argument Davidson gives against the name theory; *op. cit.*, p. 83.

To see how quotation is supposed to run afoul of the principle, we need only consider definitions. There is a well-established practice of using quotation to introduce expressions in definitions. (12) is typical:

12. '⌖' means *alternatively*.

The argument assumes that semantic values of basic expressions which can occur in sentences of a language must have been given by stipulation. The quoted expression in (12) has a semantic value (something like a shape), which determines the content of the definition, what the definition is a definition of. The expression is also clearly simple. The sheer variety of expressions that can be used in this way makes it impossible for such expressions to have been constructed from a finite number of more basic parts. Therefore, the expression in (12) and others like it must have had their values stipulated. But the number of expressions which can be used in this way is infinite. Admitting such expressions into defining sentences will violate the Davidson's finiteness constraint.

I cannot say that Davidson offers a solution to this problem, since he does not state the argument. One aspect of his theory, however, looks very much like a response to difficulties raised by the argument. Davidson excludes quoted expressions from quoting sentences. He admits into sentences only expressions which occur outside of quotation, expressions which have already been defined. Because only a finite number of definitions will have been introduced at any given time, the number of expressions able to occur in sentences will be finite, at any given time. Finiteness is preserved.

I think that such drastic measures are unnecessary and that the argument that motivates them is unsound. The weak point is the assumption that the semantic value of a semantically primitive expression must have been stipulated. This assumption rests on the view that that stipulation is the only way in which a semantically simple expression can acquire a semantic value, something which is true if all uses of expressions are standard uses, but is not true if, as the identity theory maintains, the semantic values of expressions in quotation are determined by a general principle. In section VI, I give arguments for why I think that the identity theory offers a correct account of how quoted expressions get their values.[19]

---

[19] At this point, I am already in a position to note that the identity theory saves the argument from infinite regress. The argument is based on the case of the

Once it is granted that the values of expressions in quotation are determined by the principle of identity, there is no longer a need to limit their number. The principle of identity, being single and finitely stateable, offers a finite interpretation for an infinite number of expressions. The constraints on finiteness only serve to limit the number of expressions in the lexicon, the set of expressions that can occur in sentences *outside* of quotation. The upshot is that finiteness constraints provide no grounds for excluding quoted material from the sentence. Therefore, they give us no reason not to take the quoted expression to be the mentioning expression.

V

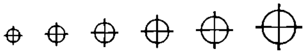
What type of object is mentioned in quotation? I believe that quotations may mention any type of expression which can be used. These expressions may be words or phrases. They may have a definite set of linguistic characteristics or they may be ambiguous. Quotation may also be used to discuss words and phrases which lack meaning, but which seem to have syntactic properties, as one finds, for example, in Lewis Carroll's "Jabberwocky." It may even be used to talk about expressions utterly devoid of linguistic characteristics. By stretching the device to its limit, one can perhaps talk about tokens or shapes.

This position contrasts with the views of all four writers both in its plurality and its specifics. These writers share a liking for uniformity. Davidson's position is distinctive in its austerity: in all of its uses, quotation mentions shapes. One of Geach's arguments suggests a position that is distinctive on the other extreme. It seems to imply that quotations always mention expressions with definite and articulate linguistic characteristics. Two other answers see all quotations as concatenations. Tarski and Quine's spelling theory implies that quotations mention strings of letters, while Geach's lexical theory implies that quotations mention strings of words. In this section, I attempt to show that the claim that all quotations mention a single type of object is theoretically unmotivated and factually untenable.

---

introduction of new expressions through definition. Definitions, however, are nothing but stipulations. (According to the identity theory, what they stipulate are standard semantic values.) If the new expressions which occur in definitions already have stipulated semantic values, which they contribute to defining sentences (as the argument assumes), we should wonder when these stipulations were performed. The assumption that the quoted expression in a definition received its value through definition leads directly to regress. On the identity theory, no regress occurs because no prior definition is necessary. Only the standard value is given by stipulation; the quotational value of new expressions is determined by the principle of identity.

Let us consider Davidson's theory. He claims that quotations mention shapes, but does not tell us what shapes are. We may presume, however, that shape would be the sort of thing that the inscriptions

13. 

have in common. It is something that they would have in common whether written in pencil, pen, or printed as they are. In the case of spoken tokens, shape is a metaphor for phonological form, a property shared by utterances of increasing volume but identical diction and accent.

Davidson's demonstrative analysis (*op. cit.*, p. 91) would give the sentence in (15) the gloss in (14′):

14. 'Snow is White' is a sentence.
14′. Snow is White. The expression with the shape of *that* token is a sentence.

In the analyzed version, the quoted expression is set off from the quoting sentence to underscore Davidson's contention that the quoted expression does not belong to the sentence as a grammatical constituent. Its occurrence within the sentence token is a reflection of nothing more than stylistic convention. In (14′) the italicized 'that' refers to the inscription 'Snow is White' as a demonstrative pronoun might refer on other occasions to objects salient in the context.

While not clearly mistaken, Davidson's commitment to shapes introduces what appear to be unnecessary complications. In his analysis, shapes enter into the truth conditions of sentences containing quotations only to give way to expressions that have that shape. It is not difficult to see why this is necessary. The same linguistic characteristics accrue to two tokens if the tokens belong to the same expression type. What is said in a use of a sentence such as (15) depends not on the shape of the token but on the expression:

15. 'SNOW IS WHITE' is a sentence.
15′. 'snOw is wHiTe' is a sentence.
15″. '*Snow is White*' is a sentence.

These sentences make the same claim because the tokens are instances of the same expression. That the quoted tokens differ in shape is immaterial from the semantic point of view.[20]

---

[20] Davidson's account of direct discourse is based on his account of quotation. According to him, a direct quotation of the form in (a) should be expanded as (b) p 92:

Davidson's preference for shapes is especially puzzling given that the content of what is said in the examples he uses never depends on the shape of the quoted token. Why then does he not say that quotations ordinarily mention expressions, when this would more accurately reflect the content of his proposal, not to speak of the true state of affairs in quotation?

Davidson does not explicitly defend his claim, but some considerations can be found in its favor. One is generality. Everything that can be embedded in quotation marks has a shape. This shape will always be there to be mentioned. Although it is a slight abuse of the device, one might argue that (16) would not be intelligible if quotation mentioned expressions, because a circle over perpendicular lines is not an expression:

16. '⌖' is symmetrical.

---

a *X* said "*P*"
b *X* said, using words of which *this* is a token, that *P*

In verbatim reports of another's speech, the quoted expression plays two roles. First, the words are ascribed to the speaker as something uttered, then they are ascribed to the speaker as the content of what was said. The theory predicts that the direct discourse report in (c) should be rendered as (d):

c. Hume said to his companions, "you must tell Mr. Adams to keep stronger chairs for heavy philosophers" (after picking himself up off the floor); John Burton Hill, *The Life and Correspondence of David Hume* (New York: Garland, 1983), p. 440
d. Hume said to his companions, using words with the shape of *this* token, that you must tell Mr. Adams to keep stronger chairs for heavy philosophers.

But this is incorrect. In general, what matters to quotation is the expression and not the shape. To report someone's words, we must use the same words as he did. (a) will be true if and only if we utter the same words that Hume uttered. Davidson's truth conditions, requiring as they do sameness of shape, are too restrictive. Indeed, on his analysis, it is hard to see how a sentence like (c) could possibly be true. If sameness of shape were what mattered, then reporting writing in speech or speech in writing would be an impossibility—on no natural understanding of shape do written and uttered tokens have the same shape. Correct reporting in the same medium would not be ruled out but would require skills of impersonation or forgery.

Indexical elements in direct quotation also pose problems for Davidson's account. As analyzed by Davidson, Hume would be reported as directing his request to my audience rather than his own and his request would apply to the present rather than the past. This is not what Hume succeeded in doing. Although I can see no reason why it would be useful to do so, one can express the truth conditions of direct quotation in terms of indirect quotation as follows:

'*X* said "*P*" ' as uttered by me now is true iff *X* said, using words of which '*P*' is a token, that *Q*.

where '*Q*' and '*P*' stand for sentences and '*Q*' as uttered by me now expresses the content of '*P*' as uttered by *X*.

(The fact that indexicals pose a problem for Davidson's account was pointed out to me by John Perry.)

That is to say, one might believe the only expressions there are are meaningful ones. On the shape theory, the problem of a lack of an available object would not arise, since the symbol above obviously has a shape. Similarly, if per chance one did want to distinguish among the quotations in (15), one might think that it would be possible to do so if quotations mention shapes, but not if they mention expressions. The thought is that shapes cut finer distinctions than expressions. If we could give correct truth conditions for all quoting sentences in terms of shapes, then shapes might form the basis of an adequate analysis.

The problem is that it does not seem possible to explain the truth of all conceivable uses of quotation using shapes. For if we are willing to accept the likes of (16), I cannot see how we could oppose

17. '⌖' is large, but '⌖' is not large.

Such a use could not be explained on the assumption that quotations mention shapes, since both occurrences of the expression have the same shape. If the search is for the greatest generality, shapes will not do. The search would inevitably lead to the conclusion that quotations mention tokens, or better yet tokenings, objects that are always present in quotation and allow for the finest distinctions.

The shape and token proposals rest on a number of questionable assumptions. One is that the uses of quotation in (16) and (17) are worthy of explanation; another is that quotations mention the same type of object in all cases. I shall not take a position on the first, but note that these uses are, at the very least, highly unusual.[21] The second, however, is clearly dubious, and as a general policy it should be resisted. If, for example, we assumed that reference to material objects is reference to a single type of entity, we would have to say that whenever we refer, we refer to space-time slices, since four-dimensional volume is the one property that all material objects have in common. The absurdity of this view reflects badly on its analogue in the case of quotation. Furthermore, I see no reason why we should be committed, as this line of reasoning seems to be, to a rigid concept of expression which requires expressions to be meaningful and which precludes any distinction between the quotations in (16). The notion of expression is flexible, and can be applied widely, making coarser or finer distinctions.

---

[21] Hans Reichenbach, in *Introduction to Symbolic Logic* (New York: MacMillan, 1947), introduces special quotation marks, which he calls "token quote-marks," which allow a speaker to refer to any object standing between them. Token quotes turn the written page into a relief map that may contain, in addition to words and punctuation, sand, home appliances, or real estate, indeed anything a person might want to talk about. For my purposes, I need only note that Reichenbach introduces special symbols to handle token mentioning quotations, rather than treat it as a special use of the familiar device

A second plausible motivation for the shape theory is related to Davidson's decision, discussed above, to exclude quoted material from the quoting sentence. He makes the connection in the following remark:

> . . . what I propose is that those words within quotation marks are not, from a semantical point of view, part of the sentence at all. It is in fact confusing to speak of them as words. What appears in quotation marks is an *inscription*, not a shape, and what we need it for is to refer to its shape (*op. cit.*, p. 90)

Davidson does not explain these remarks, but it seems reasonable to think that his belief that quoted expressions are not part of quoting sentences led him to conclude that quoted expressions are not words; where by 'words', I suppose, he means *interpreted linguistic expressions*. The general idea would be that tokens that are not used do not possess linguistic characteristics. Quoted tokens would therefore be mere inscriptions. Since inscriptions have no linguistically relevant properties aside from their shape, it follows that, if quotations mention what stands between the marks, quotations must refer to either the inscription itself or its shape. We should note that while Davidson's stated position is that quotations mention shapes, his treatment is compatible with the assumption that they mention tokens.

This argument assumes that quoted expressions are not part of quoting sentences. We have observed that, if the identity theory is adopted, there is no reason to exclude the quoted material from the quoting sentence. With the material readmitted to the sentence as an expression, it can serve as the semantic value of the quotation. Provided that the value-fixing mechanism of the identity theory can be defended, it follows that the second reason for saying that quotations mention shapes is no good.

Let us now turn to the description theories, which are another source of uniformity. On the spelling theory, each multiletter quotation is short for a description that describes an expression in terms of its constituent letters. The nature of the descriptions implies that all quotations mention series of letters. The main problem with this consequence is that it cannot be reconciled with conceptions of syntax on which the syntactic resources of a language are more advanced than concatenation. Most current syntactic theories recognize constituents and subconstituents, whose relationships cannot be represented in the linear ordering of letters. If linear ordering were all there is to the values of quotations, then most of what linguistics say about expressions could be prejudged to be false. The lexical description theory has consequences that are just as intolerable.

It is interesting to note that the considerations Geach cites in favor of his theory actually support the idea that quotations mention expressions whose structures are not linear. The relevant comments come in the following passage:

> a quoted series of expressions is always a series of quoted expressions; the quotes around a complex expression are to be read as applying to each syntactically distinct part of the expression; and if we fail so to read them, because we cannot discern the syntactical parts of the quoted expression, then we likewise fail to understand the quotation—we do not know what it is a quotation *of* (*op. cit.*, p. 82).[22]

In line with his official position, Geach emphasizes the serial order of constituent expressions. Yet if he is concerned, as he says he is, with the syntactic parts of sentences, he must be prepared to recognize syntactic relations that are more articulate than concatenation and constituents that are larger than words. If quotes apply to each syntactically distinct part of an expression, they must apply to noun phrases as well as the nouns they contain, verb phrases as well as the verbs they contain, etc. But this means that quotations must abbreviate descriptions of entire syntactic structures. Given that quotations must correspond word for word with the descriptions they abbreviate, it follows that they must describe single syntactic structures. For in order for a quotation to be able to mention more than one structure, it must have constituents that can refer to the same word on more than one occasion. But a single syntactically determinate expression is an unambiguous expression. Geach's argument suggests, therefore, that quotations mention syntactically unambiguous expressions.

This high degree of specificity has unfortunate consequences. The fact that I say something true when I quote the expression in (18) and say of it that it is ambiguous shows that quotations need not mention syntactically unambiguous expressions.

18. Flying planes can be dangerous.

Geach suggests at other points in his discussion that quotations mention words bound to particular languages (*op. cit.*, p. 86). If this were so, then it would not be possible to mention an expression common to distinct languages. The fact that I say something true when I quote the expression in (19) and say of it that it is a noun in English and a preposition in French shows, however, that quotation can mention language-neutral expressions.

---

[22] This passage is slightly ambiguous, since it is not clear whether quotations are "of" what is used or what is mentioned. But this ambiguity will not prevent us from evaluating the proposal.

19. Car[23]

Since quotation can always be used to mention the whole range of unambiguous expressions, these facts suggest the quotation can be used to mention any type of linguistic expression.[24] Of course, this is

---

[23] I assume that semantics is determined at surface structure. If the interpretation is determined at deep structure and (21) is derived by a "conjunction reduction" transformation from 21′, then the semantics could assign distinct references to the different occurrences of "car":

21′. 'Car' is a preposition in French and 'car' is a noun in English

There are no good arguments for this hypothesis, however.

[24] Up to this time, I have spoken freely of types, but it should be clear that they are nothing more than a device to aid in facilitating generalizations, a benefit which can be discarded at any time. Indeed, it is a reasonable question to ask what an interpreted or uninterpreted type is. This question can be answered by describing the relation between types and tokens of those types (For a discussion of the significance of the role of types in linguistic theory, see Sylvain Bromberger's "Types and Tokens in Linguistics," CSLI-report 88-0125.)

Let us consider what I have called "uninterpreted expressions " I have discussed the use of such expressions in definitions like (12) I might have carried out this definition to the same effect using a different token, such as the one in (12′):

12′. '⊕' means *alternatively*.

What is important to the definition is not the particular token but its characteristics. Following Bennett, I call characteristics of a token which matter to interpretation the token's "linguistically relevant characteristics." For the purposes of language use, there is no choosing between tokens with the same linguistically relevant characteristics. When we give definitions, the linguistically relevant characteristics will be some subset of the token's physical characteristics. For written tokens, these will include form and perhaps size, but not weight or material composition. For spoken tokens, accent will be important, but volume rarely, if at all. When a token is used in a definition, it is an example that serves as a paradigm for future uses. (It is important to underscore that this is true only of expressions that contain no linguistically interpreted parts. For example, if the new expression contains letters, then these letters will not be characterized by the form of the letter tokens that are produced in the definition.)

When it comes to interpreted expressions, the situation is more involved Since tokens that look or sound alike may have different linguistic properties, the linguistically relevant properties of interpreted tokens are not limited to physical properties. In addition, the intentions of the speaker are also relevant. Through these intentions, the token is associated with a certain standard usage. This standard usage determines its standard semantic value The speaker intends to use the token as others have used similar tokens in the past. A speaker will not be interested in associating his use with just any previous uses. The only uses which matter are those connected to the current use by a chain of communication Uses completely cut off from the current one will not be of interest (The picture is implicit not surprising. The variety of mentioned expressions simply reflects the variety of used expressions; a quoted expression mentions itself.

VI

How is the semantic value of the mentioning expression determined? We have decided that in quotation the expression used and the expression mentioned are one and the same. The question that remains is how the expression is associated with itself. The name, description, and identity theories all offer different answers to this question.

The name theory disregards all appearances of structure. It sees all quotations, whether of a word, a sentence, or an entire discourse, as single semantically simple expressions. The description theories, which recognize structure where the name theory does not, select quotations of words or letters as semantically simple and builds complex quotations from them. What the theories have in common is their reliance on standard semantics and semantically simple expressions, though their selections of simple expressions differ. The theories recognize only the standard use of simple expressions. Now, the defining characteristic of a semantically simple expression is that its standard semantic value is determined by stipulation. I shall argue that this feature of simple expressions is incompatible with the facts of quotation. Neither whole quotations nor their parts can receive their semantic values by stipulation. By approaching the issue from different though related perspectives, I hope to highlight the consequences of the differences between the name and description theories, on the one hand, and the identity theory, on the other.

---

in Saul Kripke and Gareth Evans, *Naming and Necessity* (Cambridge: Harvard, 1980) and "The Causal Theory of Names," in *The Philosophy of Language*. Physical form enters the picture insofar as the attributes of a token cannot deviate arbitrarily far from some paradigm and still be identified with other tokens.

This picture applies not only to names and natural-kind terms, but to all basic expressions. Even the use of a demonstrative involves an association with a network of past usage. This is what makes a use of the token 'that' a use of a demonstrative. The difference between indexicals, like pronouns and demonstratives, and nonindexicals, like names, predicates, and conjunctions, is that once the latter are connected with a standard use their semantic value is determined, whereas the standard use only determines the rules of reference of the former. These rules set conditions under which the particular token picks out a particular object. (Similar views can be found in D. Kaplan, "Words," *Proceedings of the Aristotelian Society*, LXIV (1990): 92–119.)

Statements about types can be understood as follows: to say that a type is used is to say that a token with certain characteristics is used. To say something about a type is to say something about tokens that have or might have had certain characteristics. We can identity an uninterpreted expression with a set of physical characteristics and a basic interpreted expression with a network of interpreted tokens.

To understand what the various views involve, it is helpful to consider the grammar of a small language containing quotation. The grammar that would result if the name theory were adopted is given below.

Lexicon			Phrase Structure
NPs	*PRED*	*ADJ*	S → NP is PRED
Otto mentions **Otto**	*big* mentions **bigness**	*very*	PRED → ADJ PRED
Otto mentions *Otto*			
big mentions *big*			
big big mentions *big big*			
and so on . . .			
very mentions *very*			
very very mentions *very very*			
and so on. .			
very big mentions *very big*			
Otto is big mentions *Otto is big*			
Otto is very big mentions *Otto is very big*			
Otto is very very big mentions *Otto is very very big*			
and so on . . .			

CHART C

Expressions in italics stand for expressions. Expressions in bold face stand for nonlinguistic entities. Quotation takes place when an expression mentions an expression. On the name theory, every unquoted expression, whether complex or simple, whether grammatical or ungrammatical, has as a counterpart a quoted expression that is a distinct semantically simple element. Every one of the infinity of quoted expressions is part of the lexicon. The grammar of the lexical description theory is the same except that only the quoted counterparts of unquoted words are part of the lexicon. Complex quoted expressions are built up from simple quoted parts. Complex quoted expressions will be well-formed NPs even if their unquoted counterparts are ungrammatical strings of words. On the spelling theory, the only quoted expressions in the lexicon are quotations of letters. The identity theory, in contrast, does not put any quoted expressions into the lexicon. The lexicon of the language above would contain 'Otto', 'big', and 'very', and a rule of interpretation would be added: quote (α) mentions α, for all α.

When the theories are presented in this way, it is easy to see the shortcomings of the name and description theories. Quoted expressions in chart C receive their values on an individual basis with no eye to regularity. This way of proceeding does not guarantee that quoted expressions match up with their values in any particular way. Whatever correlations arise are the result of accident, not planning. So, when we look at the grammar, we have to think that it is just a coincidence that 'very' mentions *very*, 'Otto is big' mentions 'Otto is big', and in general, expressions mention themselves. It would not go against the semantic rules if (20) were to turn out to be false because 'Otto' mentioned 'supercalifragilisticexpialidocious':

20. 'Otto' is big.

This shows that the grammars of the name and description theories do not capture the phenomenon of quotation. It is of the essence of quotation that quoted expressions mention themselves and this essence is not an essential part of these theories.[25]

The productivity of quotation also poses insurmountable problems for the name and description theories. The name theory treats each quoted expression as a special case requiring individual stipulation. Since only a finite number of stipulated expressions can occur in sentences of the language, the name theory predicts that only a finite number of expressions can occur in quotation. Previous observations have already shown that this is not the case. I noted that an infinite number of new expressions may be defined using quotation and these expressions will be part of quoting sentences.[26] Indeed, there is no need to go beyond expressions that already have meaning. There are an infinite number of meaningful expressions in English and everyone of them can be quoted. If the name theory were correct, this would not be possible.

The productive resources of the description theories put them in a somewhat better position to confront the productivity of quotation. By choosing one simple quoted expression for each simple unquoted expression, the theories generate an infinite number of quoted expressions that mirror the infinite number of unquoted expressions. The problem with these theories is that not all expressions can be built out of the simples currently in the language. For instance, all potential new expressions cannot be built out of a finite number of parts. Since descriptions must be constructed from a finite basis and any expression may occur in quotation, there will not be enough descriptions to describe all expressions that may occur in quotation. For this reason, the descriptions theories cannot be correct as general accounts of quotation.[27]

---

[25] This objection to the name and description theories is in the spirit of those offered by Davidson; see *op. cit.*, pp. 83, 87

[26] This is essentially the line taken by the argument in section III That argument treated the occurrence of new expressions as the standard occurrence of simple expressions

The third and last aspect of quotation that I shall consider is its opacity. The opacity of the quotation is illustrated by the fact that substitution of 'Clemens' for 'Twain', which in other cases would be licensed by (23), leads from the true (22) to the false (24):

22. 'Twain was a writer' has 15 letters.
24. Twain = Clemens
24. 'Clemens was a writer' has 15 letters.
25. Twain was a writer
26. Clemens was a writer.
27. Clemens
28 Twain

The truth of (22) and the falsehood of (24) shows that the names 'Twain' and 'Clemens' cannot be used interchangably in quotation. The name and description theories explain this fact in two stages. First, they determine the semantic structure of the quoted expression, and then assign semantic values to the simple parts. The name theory treats the entire quoted expressions in (22) and (24) as simple wholes. As simples, it assigns to them the expressions in (25) and (26) as semantic values. The difference in the lengths of the semantic values of the quoted expressions explains the difference in the truth values of (22) and (24). The fact that the unquoted sentences have the same semantic value is irrelevant. The lexical description theory assigns the quoted expressions a structure in which the words are semantically simple. It assigns the expressions in (27) and (28) to the quoted names 'Twain' and 'Clemens'. The difference in the lengths of the semantic values of the quoted names makes a difference in the length of the semantic value of the quoted sentences, which explains the difference in truth value between (22) and (24). The fact that the unquoted names have the same semantic value is irrelevant. Similar things are true of the spelling theory.

One should not be impressed by the apparent success of these theories in explaining opacity. Their emptiness is obvious. The weakness is again due to the stipulative nature of the semantics. In each case, the difference in truth value comes down to the difference in the semantic value of the basic expressions, values which are perfectly arbitrary. Without some principle backing the assignment of values, the explanation of the difference in the truth values is simply that quoted sentences mention different things. But this much we knew before we got started.

---

[27] These objections are again essentially Davidson's; see *op. cit.*, pp. 83, 89

The identity theory furnishes simple and natural explanations for the facts that cause trouble for the name and description theories. First, consider the fact that a quoted expression invariably mentions itself. This correlation appeared on the name and description theories to be merely a fortuitous coincidence. According to the identity theory, it is the rule of quotation for which there are no exceptions. If an expression does not mention itself, then it is not quoted. Second, the fact that we can quote an infinite number of meaningless expressions, which turns out to be incomprehensible on the name and description theories, is not an issue for the identity theory. Using a single principle, the identity theory churns out interpretations of all quotations. The principle applies with equal indifference to interpreted and uninterpreted expressions, and applies to a finite number just as quickly as it does to an infinite number. As a result, problems having to do with the size of the interpretative apparatus do not afflict the identity theory. Finally, there is the phenomenon of opacity. Using the distinction between quotational and standard uses of expressions, we can say that the use of an expression is opaque if its standard semantic value may change when one of its constituents is replaced by another expression with the same standard semantic value. By tying the semantic value of a quoted expression to the identity of the expression, the identity theory makes it clear why uses of expressions in quotation are opaque. In the case discussed above, replacing 'Twain' by 'Clemens' in (22) lengthens the expressions used and so lengthens what is mentioned, making (24) false. The fact that 'Twain' and 'Clemens' have the same standard semantic value is irrelevant.

Although Davidson's demonstrative theory is false, I believe that quotation has more in common with demonstration than with naming or describing. We can demonstrate an object through one of its tokens. I can mention a type of car by pointing to a particular car and saying: "*that* model" or "the model of *that* car." In demonstration, the only requirement is that the object or its proxy be salient in the context. In quotation, we also mention an object through one of its tokens. The difference between quotation and demonstration is that quotation makes a more economical use of resources. Since what is mentioned is a piece of language, we economize and use it in a syntactic and semantic capacity in the sentence.

COREY WASHINGTON

University of Washington

NOTES AND NEWS

The Association for Practical and Professional Ethics announces its second annual meeting, to be held March 4-6, 1993, at the University of Maryland/College Park. The program will include: a keynote address by Sissela Bok; papers and presentations on ethical concerns in various fields; papers and presentations on ethical issues that cut across disciplines; demonstrations on the pedagogy of teaching ethics. For more information, contact Brian Schrag, Executive Secretary, Association for Practical and Professional Ethics, 410 North Park Avenue, Bloomington, IN 47405.

Jones and Bartlett Publishers is pleased to announce that it has activated its textbook publishing program in philosophy, under the direction of Arthur C. Bartlett, and that it has acquired the distinguished philosophy list of Freeman, Cooper, and Company. Under the Consulting Editorship of Robert Ginsberg of Penn State University, Jones and Bartlett is actively seeking philosophy texts in traditional fields, as well as in Feminist Theory, Applied Ethics, African Thought, Philosophy of Aging, Health Care, etc. Requests for examination copies or proposals for new volumes may be addressed to Arthur C. Bartlett, Jones and Bartlett Publishers, Inc., 30 Granada Court, Portola Valley, CA 94028-7730.

Rodopi Publications is pleased to announce a program of scholarly works in philosophy, the *Value Inquiry* Book Series, under the Executive Editorship of Robert Ginsberg, Penn State University. Monographs, thematic volumes, special series, conference proceedings, and translations into English will be considered in all areas of value inquiry. For further information, contact Robert Ginsberg, Executive Editor, Value Inquiry Book Series, Penn State Univ., Delaware County Campus, Media, PA 19063-5596

Routledge is pleased to announce the publication of *The International Journal of Philosophical Studies* (formerly *Philosophical Studies,* Ireland) beginning with Volume I, Number 1, March 1993. The Editor is Dermot Moran, and the Reviews Editor is Maria Baghramian. The aim of the journal is to publish articles of the highest academic quality in all areas of philosophy. In particular, it seeks to promote dialogue between "analytic" and "continental" styles of philosophy, as well as to understand how issues in the history of philosophy shed light on current debates. There will be a comprehensive Book Review section, as well as a Critical Notice section where a major book (or theme bringing together several books) will be examined in greater depth. For information on submissions, write to Dermot Moran, Editor, The International Journal of Philosophy, Phil. Dept., University College Dublin, Dublin 4, Ireland, on book reviews and critical notices, to Maria Baghramian (also University College Dublin); on subscriptions, to Trevina White, Routledge Subscriptions, ITPS Ltd., Dept. J, Cheriton House, North Way, Andover, Hants SP10 5BE, England.

Chicago

Oneself as Another

Paul Ricoeur

Translated by Kathleen Blamey

Ricoeur's greatest work to date, *Oneself as Another* focuses on a major theme of contemporary philosophy: the concept of personal identity. Ricoeur's primary aim is to develop a hermeneutics of the self that will chart its epistemological path and ontological status. A second aim is to differentiate two meanings of identity that will give rise to a dialectic of selfhood and sameness.

Cloth $29.95 384 pages

Notebooks for an Ethics

Jean-Paul Sartre

Translated by David Pellauer

Presented here for the first time in English, *Notebooks for an Ethics* is Sartre's attempt to develop an ethics consistent with the profound individualism of his existential philosophy. He grapples anew here with such central issues as "authenticity" and the relation of alienation and freedom to moral values.

"A classic deserving equal status with *Being and Nothingness* and *Critique of Dialectical Reason*." —Annette Lavers, reviewing the French edition in *Times Literary Supplement*

Cloth $49.95 608 pages

Truth and Existence

Jean-Paul Sartre

Translated by Adrian van den Hoven

Edited and with an Introduction by Ronald Aronson

At long last, *Truth and Existence* allows us to read Sartre's analysis of knowing and truth. Written in 1948 and published in France in 1989, this brief, coherent, and engaging text presents Sartre's ontology of truth in terms of his characteristic key moral questions of freedom, action, and bad faith. Here is Sartre the existentialist at his most original and most provocative.

Cloth $18.95 144 pages

The University of Chicago Press

Chicago

Phenomenology and Deconstruction, Volume II

Method and Imagination

Robert Denoon Cumming

In this volume, Cumming analyzes Sartre's transformation of Husserl's phenomenological method, providing an introduction to phenomenology and to the ways in which debts to previous philosophies can be refurbished in later philosophies.

Paper $21.95 392 pages (est.)
Library cloth edition $52.95
Volume I also available

Essays on Kant's Political Philosophy

Edited and with an Introduction by Howard Lloyd Williams

Thirteen distinguished contributors from the United States, Canada, Britain, and Germany cast light on important aspects of Kant's liberal thinking. Key topics covered include Kant's liberal reformism, his relation with Hegel, his attitude to women, the use of reason, revolution, Kant's optimism and his moral and legal rigorism.

Cloth $32.50 352 pages

On Intersubjectivity and Cultural Creativity

Martin Buber

Edited and with an Introduction by S. N. Eisenstadt

These sociological writings by one of the foremost religious and social philosophers of the twentieth century offer essential insights into the human condition as it is expressed in culture and society

Paper $15.95 304 pages
Library cloth edition $39.95
Heritage of Sociology series

The Philosophy of Grammar

Otto Jespersen

With a new Introduction by James D McCawley

This book is Jespersen's critical overview of the full gamut of grammatical concepts. Jespersen discusses issues whose importance has become recognized only decades after his death.

Paper $19.95 380 pages

On What We Know We Don't Know

Explanation, Theory, Linguistics, and How Questions Shape Them

Sylvain Bromberger

In this collection of essays, Bromberger explores the centrality of questions and predicaments they create in scientific research. He discusses the nature of explanation, theory, and the foundations of linguistics.

Paper $16.95 240 pages (est.)
Library cloth edition $41.00

PAPERBACKS

Socrates' Second Sailing

On Plato's Republic

Seth Benardete

"Cryptic allusions, startling paradoxes, new questions . . . about long-accepted passages all work to give brilliant new insights into the Platonic text and therewith as well the possibility or impossibility of combining politics and philosophy, practice and theory." – Arlene W. Saxonhouse, *Political Theory*

Paper $14.95 248 pages 14 figures

Cosmopolis

The Hidden Agenda of Modernity

Stephen Toulmin

"[Toulmin] has now tackled perhaps his most ambitious theme of all. . . . By charting the evolution of modernity, he hopes to show us what intellectual posture we ought to adopt as we confront the coming millennium."
– Quentin Skinner, *New York Review of Books*

Paper $13.95 240 pages

Postmodern Sophistications

Philosophy, Architecture, and Tradition

David Kolb

"Kolb's book constitutes a welcome, accessible, and reassuring contribution to the discussion of architecture and philosophy." – Margaret Soltan, *American Literary History*

Paper $13.95 230 pages 8 halftones

THE UNIVERSITY OF CHICAGO PRESS

Oxford ✦ University ✦ Press

New in paperback!

Speech, Crime, and the Uses of Language

KENT GREENAWALT

"So long as someone like Kent Greenawalt is at the scales there is no reason to fear that free speech will be balanced lightly away; Greenawalt's earnest concern for free speech shines through in every case he considers."—*Constitutional Commentary.*

1989 (paper 1992) 368 pp.
paper $19.95

New in paperback!

The Justification of Science and the Rationality of Religious Belief

MICHAEL C. BANNER

"A significant contribution to an important trend in the philosophy of religion."—*Journal of Religion.* In this critical examination of recent accounts of the nature of science and of its justification, Banner contends that models of scientific rationality that are used in criticism of religious beliefs are in fact often inadequate as accounts of the nature of science.

(Oxford Philosophical Monographs)
1990 (paper 1992) 208 pp.
paper $22.00 cloth $55.00

Human Knowledge and Human Nature

A New Introduction to an Ancient Debate

PETER CARRUTHERS

"Instructive and provocative."—*Times Literary Supplement.* Carruthers aims to transform and reinvigorate contemporary empiricism, while also providing an introduction to a range of issues in the theory of knowledge.

1992 208 pp.
paper $13.95 cloth $55.00

A Daoist Theory of Chinese Thought

A Philosophical Interpretation

CHAD HANSEN

"The time is ripe for this sort of wide-ranging reappraisal. I believe the book will be a major factor in setting the tone and parameters of the debate for the next decade."—Victor Mair, *University of Pennsylvania.* This ambitious book presents a new interpretation of Chinese thought guided both by a philosopher's sense of mystery and by a sound philosophical theory of meaning.

1992 464 pp. $65.00

New in paperback!
Winner of the 1991 Bertrand Russell Society Book Award

Russell, Idealism, and the Emergence of Analytic Philosophy

PETER HYLTON

"Here, at last, is a philosophically sophisticated, historically sensitive, and richly detailed account of the events that led to the overthrow of the sort of idealism that prevailed at Oxford and Cambridge at the turn of the century and its replacement by so-called analytic philosophy....A splendid book. A most welcome achievement and a must for any academic library."—*Choice.*

1990 (paper January 1993) 440 pp.
paper $22.00 cloth $69.00

The Common Mind

An Essay on Psychology, Society, and Politics

PHILIP PETTIT

What makes human beings intentional and thinking subjects? How does their intentionality and thought connect with their social nature and their communal experience? *The Common Mind* argues for an original way of marking off thinking subjects, in particular human beings, from other intentional systems, natural and artificial.

1992 352 pp. $39.95

Malebranche and Ideas

STEVEN M. NADLER

In this work, Nadler examines Malebranche's theory of ideas and the doctrine of the vision in God with the aim of replacing the standard interpretation of Malebranche's account with a new reading. He argues that Malebranche's ideas should be seen as essences or logical concepts, and that our apprehension of them is thus of a purely intellectual character and serves to provide us with knowledge of eternal truths.

1992 216 pp. $39.95

New in paperback!

Realism in Mathematics

PENELOPE MADDY

"Written in a lively, engaging style. We hope that it serves to stimulate others to think seriously about issues in philosophy of mathematics because, as Maddy claims, these issues bear directly on mainstream philosophy."—*Philosophy of Science.*

1990 (paper 1992) 216 pp.
paper $17.95 cloth $45.00

False Consciousness

DENISE MEYERSON

Meyerson uses the techniques of analytic philosophy to investigate Marxists' view of pervasive irrationality and argues that Marxism is committed to the idea of motivated belief. She shows that there are other philosophically defensible claims which are congenial to Marxism: that there are facts about interests, that interests are not based on wants, that a desire can be contaminated by its history, that our judgments about our interests do not automatically motivate us, and that beliefs can survive the evidence that they are false. *(Oxford Philosophical Monographs)*

1991 200 pp. $49.95

Warrant

The Current Debate

ALVIN PLANTINGA

"This two-volume work is one of the major accomplishments of twentieth-century epistemology."—Richard Foley, *Rutgers University.* Plantinga examines the nature of epistemic warrant; whatever it is that when added to true belief yields knowledge. This volume surveys current contributions to the debate and paves the way for his own positive proposal in *Warrant and Proper Function.*

1992 256 pp.
paper $19.95 cloth $39.95

Warrant and Proper Function

ALVIN PLANTINGA

In this companion volume to *Warrant: The Current Debate,* Plantinga develops an original approach to the question of epistemic warrant; that is what turns true belief into knowledge. He argues that what is crucial to warrant is the proper functioning of one's cognitive faculties in the right kind of cognitive environment.

1992 272 pp.
paper $19.95 cloth $39.95

Practices of Reason

Aristotle's *Nicomachean Ethics*

C.D.C. REEVE

In a striking reversal of current orthodoxy, Reeve argues that scientific knowledge (*episteme*) is possible in ethics, that dialectic and understanding (*nous*) play essentially the same role in ethics as in an Aristotelian science, and that the distinctive role of practical wisdom (*phronesis*) is to use the knowledge of universals provided by science, dialectic, and understanding so as to best promote happiness (*eudaimonia*) in particular circumstances and to ensure a happy life.

1992 240 pp. $49.95

Prices are subject to change and apply only in the U.S. To order, send check or money order to: Humanities Marketing, Dept. MQ. To order by phone using major credit cards please call 1-800-451-7556.

200 Madison Avenue ✦ New York, NY 10016

FOR THE BEST IN PHILOSOPHY TURN TO CAMBRIDGE

Ideals as Interests in Hobbes's *Leviathan*
The Power of Mind Over Matter
S. A. Lloyd
"For those who want to understand what Hobbes's political theory *is*, this will be *the* book to consult." —*Alan Donagan, California Institute of Technology*
39243-8 Hardcover $54.95

Human Beings
David Cockburn, Editor
What is the importance of the notion "human being"? American and European contributors eloquently defend their views in ways that appear in sharp contrast to current work in moral philosophy and philosophy of mind.
Contributors: *Raimond Gaita, Christopher Cherry, Lars Hertzberg, Paul-Francis Snowdon, John Haldane, S. R. L. Clark, David Andrew McNaughton, Cora Diamond, Geoffrey Madell, Fergus Kerr, Ilham Dilman, O Hanfling, Edward Jonathan Lowe, Kathleen V. Wilkes.*
Royal Institute of Philosophy Supplements 29
42245-0 Paper $22.95

Now in paperback...
A History of Twelfth-Century Western Philosophy
Peter Dronke, Editor
"...presents a very useful picture of the current state of scholarship in the areas it addresses."
Contributors: *Peter Dronke, Winthrop Wetherbee, Tullio Gregory, Michael Lapidge, Jean Jolivet, Charles Burnett, Karin Margareta Fredborg, Martin M. Tweedale, Klaus Jacobi, Stephen Gersh, D. E. Luscombe, Dorothy Elford, John Marenbon, Danielle Jacquart, Enzo Maccagnolo.*
25896-0 Hardcover $84.95
42907-2 Paper $27.95

Cambridge Texts in the History of Political Thought

Mill: Political Writings
Terence Ball, Editor
38323-4 Hardcover $49.95
38748-5 Paper $15.95

Bodin: On Sovereignty
Julian H. Franklin, Editor
34206-6 Hardcover $49.95
34992-3 Paper $15.95

Nicholas of Cusa
The Catholic Concordance
Paul E. Sigmund, Editor
40207-7 Hardcover $69.95

Price: Political Writings
D. O. Thomas, Editor
40162-3 Hardcover $49.95
40969-1 Paper $16.95

Harrington: *The Commonwealth of Oceana* and *A System of Politics*
J. G. A. Pocock
41189-0 Hardcover $54.95
42329-5 Paper $17.95

Available in bookstores or from

CAMBRIDGE UNIVERSITY PRESS

40 W. 20th St., NY, NY 10011-4211
Call toll-free 800-872-7423
MasterCard/VISA accepted
Prices subject to change.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NUMBER 12
DECEMBER 1992

page
607 *Knowledge of the Past and Future*
Gerald Feinberg, Shaughan Lavine, and David Albert

COMMENTS AND CRITICISM
643 *Is It Less Wrong to Harm the Vulnerable than the Secure?*
Frances Howard-Snyder

BOOK REVIEWS
648 Charles S. Chihara: *Constructibility and Mathematical Existence*
Michael D. Resnik

652 INDEX TO VOLUME LXXXIX

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Leigh S. Cauman; *Treasurer,* Barbara Gimbel; *Other Trustees,* Kent Greenawalt, Michael J. Mooney, Daniel Shapiro.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

1993

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$65.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$10 00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1992 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIX, NO. 12, DECEMBER 1992

KNOWLEDGE OF THE PAST AND FUTURE*

Most of the knowledge to which scientists aspire is of events not taking place at present. We construe the world as being composed of events that stretch from the distant past, as much as twenty billion years ago, to a time that is perhaps much further in the future. Our knowledge of individual events that are observationally remote from us is of a different character than is our knowledge of readily observed events. We cannot hope to learn about most remote events by simple observation. Instead, we depend on the processes of prediction—for the future, and retrodiction—for the past, in addition to observation, our most reliable source of information. Since prediction and retrodiction are fundamentally theoretically based processes, the ultimate scope of predictive and retrodictive knowledge depends on basic physical theory much more than does observational knowledge. As a result, our possibilities for knowledge of the past and future have been influenced by modern developments in physical science. We discuss those developments and their implications for prediction and retrodiction here.[1]

Prediction and retrodiction are often taken to be symmetric processes, and indeed we believe that they are in certain circumstances, such as those arising in celestial mechanics. Since in other circum-

---

* The Editors regret that Professor Feinberg died shortly after this jointly written article was accepted. Please see our June 1992 issue for an obituary

[1] It ought to be remarked that we shall be discussing this largely from the perspective of what have become the standard interpretations of those developments; and (more particularly) that we shall be discussing the implications of quantum mechanics for prediction and retrodiction largely from the perspective of the Copenhagen interpretation.

One of us (D.A.) has written elsewhere on these questions from very different perspectives.

0022-362X/92/8912/607–42

stances retrodiction involves additional complications, however, we begin with the more familiar case of prediction.

PREDICTION

The ability to predict future events has been one of the hallmarks of science ever since Thales predicted a solar eclipse in the sixth century B.C. Progress in astronomy and mechanics, through the work of Kepler, Newton, Laplace, and Einstein has been tied to ever more accurate predictions of the positions of the sun, moon, and planets. By the nineteenth century, it had become possible, through the use of sophisticated mathematical analysis applied to Newton's laws of motion, to predict the positions of the bodies in the solar system for hundreds of years into the future. In the twentieth century, scientists have made extremely accurate predictions of future motion, both of objects within the solar system and of certain extrasolar double-star systems, and have used these predictions to test some of the consequences of general relativity theory.

The predictions made in Newtonian mechanics are of the motion of widely separated compact objects, or "particles." While analogous predictions can be made for systems that do not consist of particles, such as electromagnetic fields as described by Maxwell's equations, the main features of the standard conceptions of prediction are already exemplified in the Newtonian case, so we shall concentrate on that at first.

I. NEWTONIAN MECHANICS

Suppose first that we wish to be able to predict future positions of a collection of particles using information that we can obtain now. The problem has three aspects: what is in principle predictable from the present physical state according to Newtonian mechanics, which of the properties of the present physical state which are necessary for such predictions are in principle discoverable according to Newtonian mechanics, and what additional limitations are imposed by the fact that we are finite creatures of bounded capacity. The third aspect is less fundamental than the other two, but at least the following three limitations that follow from it are general enough to be worth considering: we cannot measure continuously variable physical quantities with perfect accuracy; we cannot make use of infinitely much information or carry out infinitely long computations; and we must gather all of our information to one location to make use of it.

Newton's laws mathematically determine the future positions of a collection of interacting particles in terms of present properties of the system, provided that the forces acting on the particles are known at all intervening times. What the forces are depends on aspects of physics which go beyond Newtonian mechanics. There are a variety of circumstances, however, in which we are confident that we know the forces that are acting. In the case of planetary motion, for example, these forces are to a high degree of accuracy nothing but the mutual gravitational attractions between the sun, the planets, and the other bodies in the solar system. Other forces, such as the pressure exerted by the "solar wind," can easily be taken into account as small corrections. The problem of knowing what forces are acting in the general case is not so much a problem of prediction as a problem of the scope of physical knowledge and should be treated in this wider context. We discuss this problem briefly below.

The properties of a system at the present time which are required for prediction are usually referred to as the initial conditions. For the solar system, insofar as the planets can be treated as mass points, the initial conditions amount to the position and velocity of each planet at one specific time, or something mathematically equivalent to this. Because the planets move in three-dimensional space, this means that at one time six numbers must be known for each planet. Armed with this knowledge, it then becomes possible to determine the future positions of each of the planets by applying Newton's laws of motion.

The above type of analysis led to one of the great triumphs of Newtonian physics. In the nineteenth century, the procedure was applied by Leverrier and Adams to the then recently discovered planet Uranus. They found that the procedure did not correctly describe the motion of this planet and drew the conclusion that an unknown force must be acting on Uranus to cause the deviation. They attributed this force to a yet unknown planet, and by a further use of Newton's laws were able to infer the position of the new planet, which was soon observed and named Neptune. Thus the failure of one type of prediction led to a successful case of prediction on another level.

Later in the nineteenth century, a similar effort to give a precise analysis of the motion of the planet Mercury using the influence of the known bodies in the solar system also showed a deviation from the Newtonian prediction. Attempts to repeat the successful discovery of Neptune by inferring a new planet, Vulcan, failed in this case, and the explanation of the discrepancy came in the twentieth century through the discovery of what is in effect an additional non-Newtonian force on the planets, whose existence is implied by the general theory of relativity.

II THE APOTHEOSIS OF DETERMINISM

Even before the application of the Newtonian method to predict the existence of Neptune, the mathematician Laplace, one of the leaders in applying Newton's laws to planetary motion, used Newton's ideas to draw a sweeping conclusion about scientific knowledge. Laplace imagined the universe as being composed of an immense number of particles, with known forces acting between them. He then described the view of such a universe which might be taken by a demonic observer who was ingenious enough to be able to determine the exact positions and velocities of all the particles at one time. For such an observer, the future motion of all the particles would be exactly predictable. With the further assumption that more complex natural phenomena could be reduced to the motion of particles, it would then follow that Laplace's demon could predict the complete future behavior of the entire universe.[2]

Laplace's vision is the apotheosis of predictive determinism. A physical theory is deterministic in the sense of Laplace if the future values of those quantities whose behavior the theory describes can be precisely predicted on the basis of some present values. In Newtonian mechanics, the quantities whose present values are needed are the positions and velocities of particles. Six numerical values must be known for each particle, so that, if the number of particles is finite, then precise knowledge of the present values of a finite number of quantities allows the prediction of the whole future. Usually, the initial conditions are taken to be the values of coordinates and velocities of all particles at one time. These are not the only values, however, that can be used to fix the solution to the Newtonian equations of motion. It is possible, for example, to specify the positions of all objects at two different times instead, or some combination of positions and velocities at different times. Although there are special circumstances in which these alternatives will not work, such as when the motion is highly periodic, there are also cases in which one or another is much more convenient than the standard initial conditions, because it is difficult to carry out sufficient measurements at a single time.

Other deterministic theories require the measurement of an infinite number of quantities at one time in order to predict the future. In Maxwell's theory of the electromagnetic field, for example, it is necessary to know the field at every point in space at one time, as well as the position and velocity at this time of all the charged particles that affect the field, in order to predict their future values. Nevertheless, if all of these are known, then the future behavior of the fields and particles can be exactly predicted.

---

[2] The recent very careful treatment of issues connected with the topics of determinism in general and Laplacian determinism in particular by John Earman (*A Primer on Determinism* (Dordrecht: Reidel, 1986)) should be consulted for further information on these topics. The problems raised there for Laplacian determinism overlap to some extent with those to be raised here, but Earman raises other serious objections there which are not directly relevant to the present discussion

Another type of prediction is possible in Newtonian mechanics which differs in important ways from Laplacian prediction.[5] Suppose that we are told that a measurement will be made of the numerical value of a specific physical quantity, such as one component of the position of a particle, at a definite time, t, in the future. Assuming that the forces acting on this particle are known, it is possible to predict the result of this measurement by carrying out a single measurement at present instead of six measurements for each particle as would be required for Laplacian prediction. What must be measured in order to predict the future result is the present value of a specific function of position and velocity, whose form depends on the forces acting and upon the time at which the future measurement will be made. If no forces act on the particle up until the time of the measurement, then the function that must be measured at present is just a linear combination of position and velocity. It is given in this case by $X = x + pt/m$. Otherwise, the function is more complicated, although it can in principle be calculated by solving the equations of motion.

On the one hand, this type of prediction, which we have called single-measurement prediction, is more economical than Laplacian prediction since it requires less present input. On the other hand, it is less powerful in some respects. The present measurement will in general be useless for predicting the results of position measurements made at times other than t, although it does allow the prediction of a specific combination of position and velocity which will vary from one time to another. For Laplacian prediction, one set of present quantities suffices to determine the results of any future measurement whatsoever. This difference can be described by the statement that Laplacian prediction allows us to predict everything about the future, while single-measurement prediction allows us to predict any specific thing that we choose. We shall see that another important difference between Laplacian and single-measurement prediction is that single-measurement prediction can be taken over with no change to quantum mechanics, while Laplacian prediction cannot.

III PROBLEMS FOR LAPLACIAN DETERMINISM

The logic of Laplacian determinism implies that, in order to predict the future, knowledge of certain quantities is necessary and sufficient. Precise prediction requires that the initial conditions be known exactly for all of the objects that influence one another. Furthermore, all of the forces that can act over the time period for which the predictions are to be made must also be exactly known. In the time since Laplace expressed his vision of a deterministic universe, it has gradually been recognized that serious problems exist with each of these requirements for prediction.

---

[5] Feinberg, Albert, and Lavine, *Physics Letters A*, CXXXVIII (1989): 454–8.

III.1. Sensitivity of predictions to initial conditions. There will always be errors in the measurements of physical quantities which make up initial conditions. These errors derive from several sources. One is the use of instruments that can in principle measure only with limited accuracy. Such errors can in principle be made smaller by the use of higher precision instruments. Although it seems clear that no continuously variable quantity can be measured with perfect precision, it is by no means clear a priori just how precisely such a quantity actually can be measured, especially in Newtonian mechanics, which does not include a theory of the structure of matter.

Another source of measurement error is the limited availability of some physical quantity, such as energy, that is needed to carry out certain measurements. According to quantum mechanics, for example, the measurement of physical structure on very small scales requires the use of very large energies. This limits the accuracy of such measurements because of the limited amount of energy in the actual universe.

It should be also be pointed out that the specification of even a single physical quantity to arbitrary precision requires giving a real number, which contains an infinite amount of information, i.e., all of its digits. A more prudent approach to what must be specified in order to make predictions would be to find out how much can be predicted on the basis of a finite amount of information, such as the first N digits of the initial position and velocity of an object.

These considerations imply that the actual value of a physical quantity may fall anywhere within a range of values around the attributed value. If a small imprecision in our knowledge of the value of the initial conditions leads to only a small imprecision in the prediction of some future value of, for example, a position, then the prediction is said to be insensitive to initial conditions. If a small present imprecision can lead to a large future one, however, the prediction is said to be sensitive to initial conditions. Sensitivity to initial conditions therefore means that a small uncertainty in an initial value may, after a time, lead to a large uncertainty in the predicted value. The influence on prediction of imprecision in our knowledge of the initial conditions generally increases with time. That is, for predictions over short time periods, a small range of initial values will always lead to a small range of predicted values; but as the time span of prediction grows, the uncertainty of prediction for a fixed range of initial values also tends to grow, and eventually becomes large enough to render the prediction useless.

An important qualification to Laplacian determinism, first realized in the last century, is that in some cases this uncertainty of prediction increases very rapidly with time, so that even an extremely small initial uncertainty rapidly results in a complete blurring of predictions.[4] This extreme sensitivity to initial conditions does not happen for all Newtonian systems, but it is sufficiently common to raise serious questions about the usefulness of the Laplacian program even within the confines of Newtonian mechanics. If we can predict accurately for only a short time into the future even if we make measurements that are as precise as we can reasonably expect, then for the system involved the times beyond this are as inaccessible to present predictions as if the theory were not deterministic at all. The existence of this problem suggests that the difference for prediction between deterministic and nondeterministic theories may be much less sharp than one might imagine. That this problem exists for many Newtonian systems has been realized for a long time, even though it is periodically rediscovered and triumphantly proclaimed anew. Poincaré, Born, and Schrödinger are but three of the scientists who have pointed out this aspect of Newtonian mechanisms in the early twentieth century, long before its current reemergence as a part of the study of "chaos."

The fact that many Newtonian systems display extreme sensitivity to initial conditions appears to limit effectively what we can know at present about the future behavior of such systems. But before accepting this conclusion, it is worth examining whether there are any alternatives.

This problem with Laplacian prediction might be avoided if it were possible to eliminate the need for specifying initial conditions. The "initial" conditions themselves are just the values of some physical quantities at a definite time. Evidently, according to Newtonian mechanics, it is possible to determine these quantities in terms of quantities at an earlier time. If everything in the universe always satisfied deterministic laws, then the behavior of everything at all times could, according to those laws, be referred back to conditions that held at the origin of the universe. As Omar Khayyam wrote, "The first morning of creation wrote what the last dawn of reckoning shall read." This does not yet eliminate the need for initial conditions. We can imagine, however, that conditions at the origin of the universe took a particularly simple form that was much more constrained than the "initial conditions" that are usually imposed at later times. If that were the case, then instead of needing to know to high precision the values of many independent quantities at the beginning of the universe, we would only need to know a few. We could then calculate what happens afterward, à la Laplace's demon. There have been some attempts in recent years to find simple initial conditions within the framework of a quantum mechanical description of the beginning of the universe.

---

[4] For a discussion of some related issues, see the article by J Ford in *Physics Today*, XXXVI, 4 (1983), p. 40.

Alternatively, it could be that the conditions that occurred at the beginning of the universe were such that its subsequent evolution was not especially sensitive to their precise values. Those systems which display extreme sensitivity to initial conditions are often such that this sensitivity only occurs in certain regions of the space of values of all initial conditions, and that in other regions the sensitivity is much reduced.

The idea that sensitivity to initial conditions has bad consequences for prediction is based on the notion that initial conditions are arbitrarily variable, so that observation is the only way of learning about them. But it may be, for example, that for some physical systems, initial conditions are not arbitrarily variable. Simple qualitative features of the system, such as its past long-term stability, may require that the initial conditions that applied in the distant past belonged to some special set of values. If this were the case, then the prediction of the evolution of such systems might not be subject to the same restrictions as predictions for systems for which the initial conditions can vary arbitrarily.

Another way of approaching predictions that are very sensitive to initial conditions is to ask what measurements are necessary in order to carry them out. To be specific, suppose we wish to predict the value of some physical quantity at a specific future time. Then we know from the discussion above that a single measurement at the present time is sufficient. A simple generalization of this technique of single-measurement prediction is that, if we wish to predict the values that *N* distinct quantities will have at various future times, which may or may not be the same, then it is always sufficient to measure *N* quantities at the present time. This type of prediction does not show extreme sensitivity to the measured values of these *N* quantities.

A much more economical approach, Laplacian prediction, is possible for systems that permit predictions that are insensitive to initial conditions, an approach in which a fixed number of measurements allows for the accurate prediction of all future quantities. It is plausible to ask whether there is any intermediate situation between the "one-for-one" requirement of single-measurement prediction and the "one-for-all" requirement of Laplacian prediction. The answer to this question is not known to us.

It is imaginable that phenomena overly sensitive to initial conditions will ultimately be described in some different way, one that does not involve initial conditions of the type usually assumed. There is no unique way to try to understand natural phenomena, and if one method fails, scientists are usually driven to devise others. For example, a description that uses a segment of a past trajectory might allow for accurate predictions even in situations where knowing a single point on the trajectory does not. It is probably premature to conclude that the extreme sensitivity of the behavior of some Newtonian systems to initial conditions entails insuperable obstacles to accurate prediction.

III.2. Other problems for Laplacian prediction.

III.2.A. Requirements beyond initial conditions. In addition to knowing the initial conditions, prediction in Newtonian mechanics also depends on knowing certain fixed parameters of a system, such as the masses of its component particles, and the constants, such as coefficients of friction, that enter into the forces that act among them. Learning the values of these parameters for a specific system involves carrying out measurements similar to those involved in learning initial conditions. Therefore, the knowledge of these parameters is subject to the same sort of inaccuracies as is the knowledge of initial conditions. Uncertainties in masses and forces can show up as uncertainties in the prediction of the future behavior of a system. Consider, for example, a body moving under the influence of a constant electric force. The acceleration will depend on this force and on the mass of the body. If those are not known with certainty, then the acceleration will be uncertain. The mathematical expression for the position of the body at some future time will include a term proportional to the acceleration, and the magnitude of this term will grow as the square of the time, so that after a long time has elapsed even a small uncertainty in the acceleration will lead to a large uncertainty in the position.

This discussion of uncertainty assumes that the mathematical form of the forces that act is known, and that only some parameters in it are uncertain.[5] Actually, if one wants to predict the position of a body at some future time, *T*, it is necessary to know the forces acting on the body at all times between the present time and *T*. Strictly speaking, we cannot be sure what these forces will be from measurements that we can do at some fixed time earlier than *T*, unless there is some theoretical input. If we assume that the system satisfies the conservation of energy, then the force function at a specific point in space is time-independent. The force function will usually vary from point to point, however, so that as the body moves it will be subject to different forces whose values cannot be anticipated from the past history of the body. In order to anticipate what forces will act on the body, one could use some theory of the spatial variation of the force, such as that provided by Newton's law of gravitation for planetary motion. Alternatively, one could try to measure the forces that would act on a body at any point in space. This is essentially what is done in mapping magnetic lines of force by the use of iron fillings. But this procedure involves going considerably beyond the 6*N* measurements that are usually described as necessary for prediction of the motion of *N* bodies in Newtonian mechanics..

III.2.B. Unobserved bodies. In Newtonian physics, the behavior of objects in some region of space can be significantly influenced by objects that are far from the region. Ordinarily, one imagines that far away objects have no important influences on a region unless they are extremely large. Even small, far-away objects, however, have some influence on what happens through the small, long-range forces that they exert. Gravity is one example of such a force. Over a period of time, such small influences can change the behavior of systems in the same sort of way as a small change in initial conditions would. This means that in order to make precise predictions about motion over long periods of time, it is necessary to have a complete inventory of all of the distant objects that might be influencing the motion. In the case of the future motion of planets in the solar system, for example, it would be necessary to include the effects of other stars, and ultimately of other galaxies, if the predictions are to be extended sufficiently far into the future. If some of these distant influences are omitted, then the predictions will become increasingly less accurate as time goes on. Ultimately, since every mass in the universe can influence every other one at least through gravity, a precise description of the future motion of any body would have to include the effects of all other bodies.

---

[5] Ernest Nagel, for example, made a related point concerning mechanical explanation See *The Structure of Science* (New York: Harcourt, Brace, & World, 1961), pp. 166–74.

This generality about prediction needs to be fleshed out by a quantitative statement about how significant the influence of small, distant objects can be. To get some idea of the size of the effect, imagine that there is an unobserved planet the size of the earth at the distance of the nearest star. The gravitational effect of this planet on the motion of the earth would correspond to an additional acceleration of about 10^{-18} cm/s^2. Such an acceleration, if it maintained a fixed direction, would produce a displacement of the earth of about 10^{16} cm over a period of ten billion years, the expected future lifetime of the earth. That distance is 100 times the size of the earth's orbit. This suggests that, if we want ultimately to extend predictions of the earth's motion to time intervals comparable to the future lifetime of the solar system, then we will need to take into account the gravitational effects of distant bodies. While this does not introduce new problems of principle, it does seriously complicate the practical problems of prediction, since one does not even know of the existence of all the distant bodies whose influence may be relevant. One might reverse the logic and argue that inexplicable deviations from motions predicted on the basis of known influences can be attributed to specific additional influences and used as constraints to help reveal the latter. The different fates of the nineteenth-century predictions of Neptune and Vulcan give some indication of the likely success of this approach. As we wish to predict the motion of any object, such as the earth, ever more accurately over greater time periods, we shall necessarily have to include the influence of ever more distant objects. Our ability to make accurate predictions will rapidly come to be limited by this need to include in our analysis increasing numbers of objects that are increasingly difficult to observe.

IV. THE CONSEQUENCES OF RELATIVITY FOR PREDICTION

In our discussion of the effects of distant bodies on local predictions, we have implicitly assumed that the distant bodies always remain far away, and that their effects are felt instantaneously. The first of these assumptions is not always true, however, and the second is never true, as a result of special relativity theory. The special theory of relativity therefore implies novel limitations to prediction which, though they may not be as profound as those of quantum theory, nevertheless force us to modify substantially what we can expect to say accurately about the future.

IV.1. Rapidly approaching bodies. Suppose an asteroid that is now very far away is heading on a course toward the earth which will lead to a collision in 100 years. While the asteroid may have only a very small effect on what happens on earth in the short run, it will likely have an immense effect on what happens during and after the impact. If we knew that the asteroid was headed our way, then we could try to predict what its effect would be. That problem is not really different from predicting the effects of any complex future event, which may be done with more or less accuracy. If we do not know about the existence of the asteroid, however, then our predictions about what will happen on the earth after the next 100 years are likely to be seriously wrong.

In a world correctly described by Newtonian physics, information could be obtained about all the bodies that in some future time period will pass sufficiently closely to a given body to influence it strongly. A proposal to obtain just such information about asteroids that might collide with the earth in the future has been considered seriously. If this were done, then we could avoid most surprises of the type just described.

It is one of the consequences of the special theory of relativity that at any given time it is impossible to obtain complete information about what may collide with us in the future. Therefore, we can never be sure that a big surprise does not lie just ahead. In order to see why this is so, let us analyze how information about distant bodies is obtained. Such information can be obtained by one of two methods, passive or active. We could passively monitor all incoming radiation, such as light, radio waves, neutrinos, and so forth that may be reaching us from afar, and use the data to infer the future motion of the bodies emitting the radiation and, in particular, whether any of them will collide with us. This procedure amounts to what happens when a pedestrian looks at other pedestrians on a street in order to avoid colliding with them. Alternatively, we may actively emit radiation, checking whether it is reflected back by unexpected objects whose presence would then be revealed. The future motion and possible future collisions with us could be inferred as in the first method. This is what is done when a ship uses its radar to detect other ships in a dense fog.

Both methods are effective for the low relative speeds of bodies on the earth, or for known bodies within the solar system. They become less effective, however, when the bodies involved are traveling at near the speed of light. Suppose, for example, that a previously undetected high-temperature jet of gas is approaching the earth on a collision course at ½ the speed of light. At some instant of time, when the jet is one hundred ten million miles away (a distance it takes light about 600 seconds to travel), we send out a radar pulse at the speed of light, which reflects from the object and returns to tell us of its existence and of the impending collision. The pulse will return to us about 800 seconds later, only about 400 seconds before the gas jet itself hits. Until the radar pulse returned, with its warning of the collision, any predictions that we would have made about events even a short time after the collision, now only 400 seconds away, could not have been correct because of information that was not available to us. If the jet were traveling at very close to the speed of light, we could get almost no advanced warning about it, either from passive or active signals. If the approaching object is moving at the speed of light, as would be the case for the radiation from a nearby supernova, we can get no advanced warning at all.

This problem occurs because according to special relativity we cannot use signals traveling faster than the speed of light to detect distant bodies. There is no such limitation in nonrelativistic physics, and the problem does not occur in this form. By using the active method with emitted radiation of arbitrarily high speed, we can always get advanced warning of an approaching body, except for the limiting case of a body that travels toward us at infinite speed. Nevertheless, if the body is approaching at some very large finite speed, the warning may give us only a very short notice of the impending collision.

It may be imagined that an observer who has been receiving signals for some time can do a better job of predicting the future. Suppose that there is some object, S, that has already been observed at a time before the present time, t_0. If these observations are sufficiently detailed, then they can be used to predict whether S will emit something subsequent to the latest time that it has been observed, which will reach the observer just after t_0. If these prior observations imply that it will not do so, then the observer has reason to believe that no surprise awaits. For example, to the extent that astronomers are sufficiently confident of their theory of how stars undergo the massive explosions known as supernovas, they could use observations about the radiation emitted by some star to infer that the star will not become a supernova a short time after the latest observation. In other words, for a body that we have observed for some time, analysis of its past behavior can be used to predict its future behavior and makes it possible to determine whether any part of the body will eventually collide with us. It is worth remarking that such analysis introduces significant new theoretical content into the bases of our predictions.

IV.2. Horizons. Even if our theories were perfect, the above type of analysis is not sufficient to rule out all surprises. In universes with certain types of cosmological structures,[6] objects may exist which could never have been observed by someone at a point, P, until the time, t_0. These objects may nevertheless emit signals that will reach P just after t_0. If the universe were finite in extent, had existed forever, and was "flat" in the sense of general relativity, then no such objects could exist. Such objects could, however, exist in a universe whose age is finite.

Suppose, for example, that the universe is static, has an age, A, and extends to a distance, D, that is considerably larger than cA, the distance that light could have traveled since the beginning of the universe. Then any objects whose distance apart is greater than cA will have been unobservable to each other until the present. Therefore, any mutual influence that such undetectable objects might have in the future will be unpredictable until it actually occurs. In the actual universe, which is expanding, the situation is still more complicated, but most of the cosmological solutions of general relativity which have been studied as models for our own universe include "horizons" of the type just discussed, that is, horizons beyond which distant objects will not become observable until some specific future time. In a universe that contains cosmological horizons, it would appear that we cannot eliminate possible surprises about what will happen locally when the influence of presently hidden distant objects begins to be felt.

It should be noted that there may be indirect ways to detect objects that are beyond such horizons. For the case that has been most completely analyzed, that of black hole horizons, some information about what is inside the horizon can be obtained. For example, the total energy and electric charge residing inside can be learned from measurements of the gravitational field and electric field outside the horizon. Also, it has been suggested that quantum effects, such as the deviation from a thermal spectrum of the "Hawking radiation" emitted by a black hole, give additional information about the matter inside the horizon.[7] It is possible that similar effects exist for cosmological horizons, but this question has not yet been investigated. It is therefore premature to draw conclusions about the implications of horizons for prediction.

IV.3. Retarded interactions. There is a further problem in relativistic theories which occurs even for systems that we can assume isolated over the period for which we wish to make predictions. In general, in such theories, the forces that act at present on an object depend on the past and present positions and velocities of all other objects within the system which can influence its motion. The role of past positions and velocities, rather than just present ones, arises because of the finite speed of light, which implies that the forces that act are retarded, that is, they depend on where the objects were rather than where they are. This means that the equation of motion of each object is a complicated integro-differential equation, involving integrals over past times, rather than a simple differential equation of the Newtonian type. This change in the form of the equations of motion leads to a change in the information that is necessary for prediction from that of nonrelativistic mechanics, in which forces act instantaneously across any distance.[8]

---

[6] The various concepts of horizons in relativity theory have been discussed by W. Rindler, *Royal Astronomical Society Monthly Notices*, CXVI (1956), p. 663.

[7] See, e.g., T.D. Lee, *Nuclear Physics B*, CCLXIV (1985), p. 437.

In a simple case, that of two particles interacting through electromagnetic forces, the force depends on the distance between the particles at a specific time in the past, known as the retarded time, which is a function of their present distance apart and the speed of light. This implies that, if we wish to predict the motion of one particle over some interval of time beginning with the present, then we must know the position of both particles from the retarded time to the present. That is immensely more information than is needed in nonrelativistic physics to carry out the same type of prediction. Whereas in nonrelativistic mechanics we need to know conditions at a single time, in relativistic physics one seems to need to know corresponding conditions over a continuous period of time.

It might be thought that the positions over the interval between the retarded time and the present could themselves be calculated from a knowledge of the positions and velocities at the retarded time alone. This calculation, however, would require knowing the forces that acted between the particles during this earlier time interval, which in turn would require knowing the positions at some still earlier retarded-time interval, and we are led to a requirement of knowing positions and velocities at an infinite number of discrete past times in order to carry out this procedure.

It is instead possible to adopt the approach of field theory, in which the forces that act on particles depend on the present values of the fields where the particles are. If the values of fields and their first time derivatives are known everywhere at the present time, as well as the present particle coordinates and velocities, then the joint evolution equations of the particles and the fields can be used to determine their future values. This procedure also requires knowing much more than is required in nonrelativistic particle mechanics, since the values of the fields at all points in space at some fixed time involve a continuum of quantities. In addition, because of the effects of horizons discussed above, it is in general not possible to determine all of these field quantities by measurements made at a fixed time. It is not necessary to know the fields at all points in space in order to calculate the force on the particles at any one time. If we want to know the force on the particles over a period of future time, however, we must know the present values of the fields even at points where the particles will never be, because those field values will influence the fields at future times at points where the particles will be. Thus, all of these approaches to prediction imply that even in the simplest case of special relativistic mechanics we require much more input to make predictions than would be needed in nonrelativistic particle mechanics. Instead of getting a large amount of information about the future from a small amount of information about the present, we need a large amount of information in order to be able to predict anything beyond the present.

---

[8] For a discussion of some of the mathematical questions arising from equations with retarded interactions, see the article by R. D. Driver, *Physical Review*, CLXXVII (1969), p. 2051

V LIMITS TO PREDICTION IN QUANTUM MECHANICS

One of the most surprising developments in twentieth-century physics has been the recognition that the fundamental laws that describe microscopic objects, and in principle macroscopic ones as well, are such that precise prediction is in most instances impossible, even for systems that are as simple as any that we know. Consider as an example a single unstable subatomic particle, such as a muon. On the average, a newly produced muon will decay in approximately two microseconds into several other subatomic particles. The time at which a specific muon will decay is not predictable according to quantum mechanics, however, even though that time can be determined once the decay has taken place. As quantum mechanics is usually presented, there are several distinct aspects to the limits on prediction. Some of these arise from one consequence of the formalism of quantum mechanics, which is that the states that are usually prepared, such as that of a single muon, do not evolve over time into states with definite values of the quantities that we would like to predict. We shall see, however, that the same formalism implies that for any quantity that we can measure in the future, there exist states which, if not disturbed by intervening measurements, do evolve into ones with a definite value of that quantity. In most cases, however, it is not at all transparent how these states can be prepared, or what their physical interpretation is.

Prediction in quantum mechanics is also constrained by the measuring process, which can introduce uncontrollable changes in the state of a system. As a result, any predictions about a system at a future time, t_F, which are based on measurements done at an earlier time, say t_0, can be upset by measurements of incommensurable quantities that are carried out at an intervening time, t_1, such that $t_0 < t_1 < t_F$. As a result, it is generally impossible to predict beyond the next measurement on a system. These different aspects of quantum imprecision must be carefully disentangled in order to appreciate properly what can and cannot be predicted.

V.1. Unpredictability due to Heisenberg's principle. The most celebrated limit to prediction in quantum mechanics is a consequence of Heisenberg's uncertainty principle, which implies that Laplacian prediction is usually impossible in quantum mechanics. Recall that in Newtonian physics, predicting the position of a particle at an arbitrary future time requires the precise knowledge of the position and velocity at some earlier time, or some equivalent information. But Heisenberg's principle entails that no experiment can reveal this information.[9] Any measurements of position and velocity are constrained by the relation $\Delta x \Delta v > h/2\pi m$, where h is Planck's constant, and m is the particle mass. In this relation, Δx is the accuracy with which the position is being measured and Δv the accuracy with which the velocity is being measured. The reciprocal relation between these accuracies allows for approximate prediction of future positions, but not exact predictions. The inherent imprecision of predictions of position are relatively unimportant for macroscopic objects, because of the factor of m in the denominator of Heisenberg's relation. To see what quantitative limits on prediction follow from Heisenberg's principle, consider a single particle moving through otherwise empty space. If the initial conditions are determined with maximum accuracy at time 0, then at a later time, t, the minimum uncertainty in the position will be:

$$\Delta x(t)_{\min} = (ht/\pi m)^{1/2}$$

What this means is that the possible position of the particle at time t will be distributed about some average value, with a "width" given by $\Delta x(t)_{\min}$. Note that $\Delta x(t)_{\min}$ increases indefinitely with time, so that eventually, unless some further measurement intervenes, the position is completely uncertain. For an electron, this width $\Delta x(t)_{\min}$ corresponds to about 1 cm at a time of one second. In other words, the future position of an electron in empty space becomes completely unknown in a relatively short time. For a virus ($m = 10^{-15}$ gm), the uncertainty width is about 10^{-5} cm, or about one viral length, 300 seconds after the initial measurement. So to the extent that a virus could be treated as a quantum mechanical particle moving in empty space, its future position could not be predicted accurately even a few minutes into the future.

---

[9] These implications are worked out in some detail in Werner Heisenberg's *The Physical Principles of the Quantum Theory* (Chicago: University Press, 1930).

Several comments need to be made about these results. First, that what holds for position does not necessarily hold for other physical properties. The momentum of an electron in empty space, once measured, is known at all future times, unless further measurements of quantities incommensurable with momentum are performed. This is because the momentum does not change in the undisturbed evolution of an electron in empty space. Indeed, for any conserved quantity, such as the total energy of an isolated system, a single measurement is sufficient to determine its values for all future times. Here the situation is the same as in Newtonian physics. Second, we have idealized the situation by assuming that the object whose future prediction we wish to predict is in empty space. The situation is somewhat different if it is acted on by forces, since those forces may tend to restrict its motion, including the "motion" that occurs as a result of uncertainty in the initial velocity. An example of this is given below.

The situation in quantum mechanics is completely different with regard to single-measurement prediction of the next measurement to be made on a system (cf. fn.3). Such predictions can be made in precisely the same way as in Newtonian physics. This is because, according to quantum mechanics, the wave function of a particle evolves in a time-reversible manner. For example, if a free particle can be prepared at time zero to have the wave function which is the one which evolves from that of a particle with a definite position at time $-t$, except that the momentum is reversed, then at a time $+t$ this particle will have a definite position, provided that there are no other measurements on the particle between time zero and $+t$. This will be true whatever the forces that act on the particle, although the actual wave function that must be prepared at time zero will depend on these forces. These results imply that there is a specific measurement, of a quantity formally analogous to that required for single-measurement prediction in Newtonian mechanics, which if carried out at $t = 0$ will allow the position of the particle at the specific future time $+t$ to be predicted with certainty. Of course, just as in Newtonian mechanics, the measurement done at $t = 0$ will not suffice for the prediction of the position at any time other than $+t$, although it will allow for the prediction of some function of position and momentum at various future times. Furthermore, the particle will in general never have a definite momentum, so that even in this case the prediction is much more limited than in Newtonian physics. Finally, for single-measurement prediction in quantum mechanics, there remains the question as to whether the function of position and momentum which must be measured at $t = 0$ in order to predict the position of the particle at time $+t$ actually can be measured. With these very important qualifications, quantum mechanics may also be thought of as a deterministic theory, in that by suitable measurements now, it is possible to predict with certainty the result of any specific measurement that is to be done at a predetermined time in the future.

Another case in which exact prediction of future position is possible is when certain forces act on a particle. For example, for a particle bound by a quadratic potential, the so-called "harmonic oscillator," if it is prepared to have a definite position at time zero, then it will also have a definite position at any later time that is a multiple of the basic period, T, of the harmonic oscillator. Thus, if we measure the position as $x = x_0$ at $t = 0$ and allow it to evolve by the equations of motion, we can safely predict that at the future time T, the position will again be x_0. Again, this prediction is only possible at specific future times, and does not hold for momentum. Furthermore, the uncertainty in position will not increase indefinitely in time in this case, but instead oscillates between definite limits.

These forms of quantum determinism do not appear to avoid the problem of limited knowledge of the future. Part of what we mean by unlimited knowledge of the future is to be able to anticipate what will actually happen. If we do not know now what measurements will actually be done some day, we are unable to use single-measurement prediction to anticipate their results. Moreover, in most cases single-measurement prediction only allows us to predict the result of the next measurement on a system, not those of subsequent members of a series of measurements. In the standard, collapse interpretation of quantum mechanics, no present measurements will allow us to predict precisely the result of future measurements that are separated from the present situation by other measurements of incommensurable quantities.

V.2. Unpredictability in particle decay. Another restriction concerning predictability contained in quantum mechanics is best typified by the decay of an unstable state. At first glance, this might also be thought of as due to a lack of knowledge of initial conditions. A uranium nucleus that decays by alpha-particle emission is a complex structure containing many neutrons and protons, and one could imagine that, if the precise motions of these particles were known, the time at which the nucleus will decay could be predicted. Of course, Heisenberg's principle forbids us from doing this, but we might try to use single-measurement prediction, for example, to infer whether the particles that are now within the nucleus will be outside of it at some specific future time. While this can be done in principle, it does not really satisfy what we would ordinarily think of as a prediction about whether the nucleus will decay. The reason is that there is a high probability that the nucleus will be disintegrated simply as a result of the measurement needed to predict the future position of the particles that are now in the nucleus. While quantum mechanics does not forbid us from carrying out this measurement, it implies that the measurement will transform the bound collection of particles into a new state that will not resemble the nucleus very much. It is this new state that will evolve in such a way that we can predict the future positions of the particles in it, not the original bound nucleus. Hence the prediction allowed by quantum mechanics is not the one that we would have liked to be able to make.

There are other circumstances in which decays appear to be unpredictable in an even more fundamental way than for an unstable nucleus, where it seems to be the initial conditions that forbid the prediction. Some unstable systems are not complex in the sense that a nucleus is. A single subatomic particle, such as a muon, is as simple a system as is known to us. Nevertheless, its decay follows the same statistical law as does the "complex" alpha decay, and the decay of an individual muon appears to be as unpredictable as that of an individual uranium nucleus. The decay of such a particle is described by an extension of quantum theory to fields. The decay can be thought of as a transition between two different states of coupled quantum fields. This transition takes place according to the Schrödinger equation applied to the coupled fields. According to that equation, the state of the entire system is a superposition of two possibilities. In one of them, the system remains in its initial configuration, corresponding to the presence of the unstable particle. In the other, the system is in the configuration in which the unstable particle is gone, and its decay products are present. The relative probability of these configurations changes with time. In the usual description, the probability of the first configuration starts out as one and predominates at early times, while that of the second starts at zero and predominates at later times. In that description, the ratio of probabilities changes smoothly with time. The probability that the system remains in the undecayed state decreases exponentially with time, and there is no sharp instant corresponding to the decay. It is therefore impossible to predict the precise time at which the transition takes place. Instead, if the system is monitored by some external measuring instrument, there is a time-varying probability that this instrument will record that the decay has occurred. Typically, there is a substantial probability of decay over some range of times which is approximately equal to the mean lifetime of the decaying state. In other words, if a particle decays with a mean lifetime of two microseconds, then a given example of that particle has a reasonable probability of decay anywhere between about zero and four microseconds after it is produced. This situation is very similar to a measurement of position in a state that is a superposition of states with definite position, where we can know in advance only the relative probability of different results.

It might appear that there is a significant distinction between the two cases, because decays take place spontaneously while position measurements require outside intervention. Actually, the distinction between the two cases is not so absolute. The determination that a decay has taken place also requires a measurement, for example, one that detects the decay products. In ordinary circumstances, these measurements are made intermittently, on a time scale that is long in comparison with the internal dynamics of the decay process. In that case, we can think of the measurement as determining, at a series of times, whether the system has decayed, much as we might measure the position of a particle at a series of times. In between these times of measurement, the system evolves without reference to the measuring process. But just as for position measurement, the determination of whether the decay has taken place does affect the system, and its subsequent evolution occurs under a new set of initial conditions. Nevertheless, the result of this process is the prediction that the probability that the initial system is intact decreases exponentially with the time that has elapsed from its original production. If the measuring process takes place on a time scale comparable to that of the dynamics of the decay process, it may result in the transfer of large enough amounts of energy to or from the decaying system to influence the decay process. In that case, there may not be an exponential time dependence for the decay.[10] But monitoring the system over such a short time scale is difficult to accomplish for subatomic particle decays.

It is not only for decays that the problem of predicting when transitions take place exists, although they illustrate it most graphically. Actually, any transition from one quantum mechanical state to another, whether it takes place spontaneously, as in a decay, or as the result of some outside perturbation, can be predicted only statistically, except in very special circumstances, as when there is a very large outside perturbation. The reason for this is the law that governs how any system governed by quantum mechanics evolves in time. According to that law, if we express the initial state as a superposition of states with definite total energy, then each term in the superposition will change over time through being multiplied by a time-dependent complex number of magnitude one. This later superposition can be reexpressed in terms of whatever states correspond to the measuring process, such as an undecayed particle and its decay products. The coefficients in this expansion in terms of the states to be measured, which depend on time, then represent the probability of transitions. Usually, these coefficients are different from zero and one, which means that the probability of a specific transition is not zero or one, but some intermediate value. As a result, the transition does not take place with certainty at a specific time, but may take place at any time for which the expansion coefficient does not vanish. The occurrence of a decay at a specific time is, according to this analysis, not very different from the actualization of a specific position for a particle as the result of a position measurement on a state that contains a superposition of different positions.

V.3. The significance of unpredictability in quantum mechanics. It is undeniable that according to the existing form of quantum mechanics, future events cannot be precisely predicted. In the standard formulation, this is simply a brute fact about the world. According to this formulation, the notion of previous generations of physicists —that the future is predictable—was simply an overfacile generalization, based on the approximate description of large-scale phenomena that we know as Newtonian physics. While we might wish that we lived in a predictable world, we have no choice but to accept nature as it is, and live with the consequences.

---

[10] A review of the literature on this effect can be found in Anthony Sudbery, *Quantum Mechanics and the Particles of Nature* (New York· Cambridge, 1986)

Alternative formulations of quantum mechanics, such as that due to DeBroglie and Bohm,[11] do not change the conclusion that the future cannot be predicted, although they modify the reasons for it. In some of these theories, the measurement process does not itself lead to additional constraints on prediction, and all unpredictability is the result of insufficient knowledge of initial conditions. In some formulations, this lack of knowledge seems to be similar to that in classical statistical mechanics, in that it describes the world as being in one of a large number of similar configurations, together called an ensemble, in which initial conditions vary from one to another. If we could somehow determine which member of the ensemble we were in, we could then make exact predictions about the future. But these alternative formulations contain no clue as to how this might be done, even in principle. Indeed, it can be proven that in order to do so, something would have to be changed about the Schrödinger equation, which describes how the quantum mechanical wave function changes with time. Theories that do this have occasionally been put forward, though not always in order to make the future predictable. If such a change were made, we would be dealing with a new theory, rather than another formulation of the same theory. Presumably, this new theory would imply that there were some additional measurements which could be made which would enable us to make precise predictions about some future event, such as the decay of an individual subatomic particle. There is no clue in anything we now know about what these additional types of measurement could be.

RETRODICTION

A large number of the events that will ever happen have already happened and we cannot plan to observe these events as we can for future events. An important question for human knowledge concerns the extent to which we can recover information about past events on the basis of present and future measurements.

While the limits for prediction which occur in Newtonian and quantum theory have been the subject of a great deal of analysis, there has been relatively little analysis of the corresponding limits for retrodiction. Such limits nevertheless exist, and while they are different in origin and scope from those for prediction, they are as important for an overall evaluation of what we ultimately can or cannot know. There are two types of retrodiction which need to be considered, which we will refer to as evolutionary and nonevolutionary retrodiction. Evolutionary retrodiction is very similar to the process used for prediction, which we have discussed above. It can be of the Laplacian or of the single-measurement type. For simple mechanical systems, such as the sun and planets taken as mass points, the same equations that determine future motions in terms of present conditions also can be integrated backward in time to give the past motions that led to the present conditions. Most of the considerations involved in such retrodiction in Newtonian physics are quite similar to those relevant to prediction, although this is not the case in quantum physics, to which we return below.

---

[11] David Bohm, *Physical Review*, LXXXV (1952), pp. 166, 180

A much greater difference between prediction and retrodiction, even in the nonquantum case, comes about, however, through the existence of alternative means to recover the past by the use of records. So far as we know, there is no analogue for knowing the future which corresponds to the use of records to know the past. Usually, one thinks of this difference as being a consequence of the fact that the past has already occurred while the future has not. We believe that instead of being a consequence of this asymmetry between past and future, the existence of records defines the asymmetry. We shall refer to any procedure using such alternative methods as nonevolutionary retrodiction.

VI HOW CAN THE PAST BE RECOVERED?

Some past events leave obvious records of themselves, as John Kennedy's inaugural address did, both in the brains of those who heard it and on many television tapes. Other events seem to vanish without a trace. The glancing of a ray of sunshine off the surface of the Pacific Ocean at 9 A.M. on May 27 in the year 1,000,001,933 B.C. would seem to have left no record of itself which can be recovered by anything that we can do now or later. There are intermediate situations in which some event left a record when it occurred, but this record has not been preserved, so that it is not available for present and future knowledge. An example would be a footprint left by someone walking on a beach, which remained for a time before being erased by the incoming tide. This simple distinction between recorded and unrecorded events, however, masks some subtleties about the physical world which need to be explored before we can decide how much of the past can be recovered. Does any event really occur without leaving a recoverable record of itself? May it be that, if we were to examine the present with sufficient care, we could find in it records of all that has ever been? For example, consider as an event a person alone in a room speaking otherwise unheard words "to the four walls." After these words are uttered, they become patterns of sound, that is, ordered motions of the molecules of the air in the room. For a short time after the words are spoken, by using sensitive devices for measuring air pressures and other physical quantities we could determine the precise pattern of sound which was emitted and so learn what was said. Indeed, the ear of another person, or the microphone of a tape recorder do just that. So, for this short time, a trace does exist of the words. But now imagine waiting for a while, say five minutes, so that the sound of the words is ostensibly no longer present, and a listener or a microphone will not be able to pick up the words. What has happened? The sound waves have interacted with objects in the room and with the walls. This interaction has changed the character of the waves so that there is no longer any very orderly motion of the air molecules related to the words originally uttered. Their trace seems to have vanished. Yet this cannot be the whole story. The phenomenon of reverberation shows that in a suitably prepared enclosure the trace of a voice or other sound can last much longer than usual and still be heard or otherwise recorded.

Even if the room does not have special properties, there is reason to think that some trace remains of any sound produced in it. To see what leads us to this conclusion, let us follow the same type of analysis for retrodiction which we have done above for prediction. In a completely deterministic world, such as one governed by Newtonian mechanics, the present state of the objects in the room is precisely correlated with their state at any instant in the past. (We assume here that the room is isolated from all outside influences, an assumption which may well be invalid in practice. We return to this issue below.) For such a world, just as we can predict the future with the aid of sufficiently exact present measurements, so we can retrodict the past with the aid of these same measurements. In this world, by making sufficiently accurate measurements on the objects in the room and using the laws of evolution, we could learn what was said in the room five minutes or even ten years ago. We have called this process evolutionary retrodiction.

Now, imagine the following sequence of events. An event occurs somewhere. For a short time thereafter, the consequences of the event are well-correlated with the event and only a small number of physical objects have been influenced by it. Measurements on this small set of objects are sufficient to carry out evolutionary retrodiction. After some time has elapsed, however, the effect of the event will become more diffuse and involve a large number of objects. In this case, evolutionary retrodiction must be applied to a much wider range of physical objects in order that an analysis of this process can lead to an inference of what happened. Thus, evolutionary retrodiction becomes more difficult as more time elapses after the event took place, because a wider range of objects becomes involved in the causal chain that begins with a specific event. A related reason why the traces of past events become increasingly difficult to detect as time passes is that most systems are open rather than closed, that is, they interact with the external world. As a result, as more time elapses, the effect of any past event is diffused over a wider and wider spatial region. In order to specify an event completely by measuring its later effects, it therefore becomes necessary to observe this wider region. The moving air molecules produced when a word is spoken collide with other molecules, which in their turn collide with still others, somewhat like the child's toy consisting of a number of steel balls suspended from a common support. In order to recapture the original sound impulse after some time has passed, it is necessary to observe most or all of the molecules that have been struck by the originally moving group, or by any that have been set into motion by them. If the room in which the sound is uttered is not very carefully insulated from the outside, this set of molecules that must be observed rapidly becomes immense.

VII THERMODYNAMIC IRREVERSIBILITY AND RETRODICTION

As indicated above, for the world as described by Newtonian mechanics, it would always be possible in principle for a closed system to recover the information about past events by careful measurements of the present state of things, combined with a use of the equations of evolution. By knowing in sufficient detail how things are now, we could not only determine what the future will surely be, but also what the past must have been. One major reason why such evolutionary retrodiction is not generally applicable to our universe is the phenomenon known as thermodynamic irreversibility. This is a fancy name for something very familiar. Imagine two bodies of water, one hot and one cold. When the two are mixed, after a short time there results a single body of water at some intermediate temperature. The mixing process has resulted in what physicists call an irreversible transition. Despite the name, this does not mean that the effect of mixing cannot be undone. Obviously, by dividing the mixture in two and applying heat or other forms of energy to the parts, it is possible to reproduce the two original bodies of water at their different temperatures. What the word 'irreversible' does mean is that this reversal will not occur spontaneously in an isolated system without the need for an external energy source, in the way that the original change to a uniform temperature takes place. Most changes that occur in nature are irreversible in this sense. The significance of irreversible transitions for retrodiction is that usually when an irreversible change takes place, the resultant system cannot be correlated to a unique predecessor. A single mixed body of water at a uniform temperature of thirty degrees Centrigrade might have come about by mixing two bodies of water whose temperatures were twenty and forty, or ten and fifty; or an infinite number of other possibilities. Once the mixed water has settled down to what is called equilibrium, it is impossible just by inspection of it to determine what bodies were combined to produce it. Even though the combined system can be returned to a situation in which its two parts are at different temperatures, there is no unique way of doing this; and if we do not know its history, we cannot expect to reproduce its actual past even by applying an external energy source to it. For a reversible change, on the other hand, it is simple to find out in what state the system undergoing the change originated. It is only necessary to let the change reverse itself, or to return the environment to its original state, and the state of the system which is reached when the environment is back in its original state is also the one in which the system started. So for this case at least there is a direct correlation between irreversibility and loss of information about the past state of a system. More generally, whenever irreversible processes occur, there is a loss of information about what existed previously. It is this loss of information that is responsible in many cases for the rapid decrease in our capability to retrodict about real physical systems. The orderly motion of air molecules produced by spoken words in the environment of the room rapidly becomes converted into disorderly motions of the type we call heat. Unless some kind of recording device is present which retains a trace of the orderly molecular motions, it seems impossible to recover the sounds even in an isolated room. This is the conventional wisdom about irreversible changes, but is it really so? It is possible to find examples of systems that undergo changes which seem very much like the irreversible processes that have been discussed, which nevertheless can be treated so that the original situation is exactly restored without the need for advance knowledge of what this situation was. Is it possible for any system to make a sufficiently accurate appraisal of its present state to distinguish among the many paths that might have led up to that state?

Even within Newtonian physics, the answer seems to be "no," if the system follows the kind of chaotic behavior described previously in connection with prediction. If the Newtonian equations for a system that exhibits such deterministic chaos are reversible, then the system will exhibit a related behavior for retrodiction. Therefore, if we know, say by observation, that the system is now in some configuration—for example, described by its being in some region of its phase space—and we wish to trace the system backward in time, then we will find that the trajectories in phase space diverge rapidly into the past as well as into the future. This will imply that a short time before we have observed the system, it may have been in one of a number of configurations which are widely different from one another, and we will have no way of knowing which of these earlier configurations was the actual one. One way to say this is that, for a chaotic system, there are many roads from many starting points which can lead to the same destination; whereas for a nonchaotic system, knowing where it ends up specifies fairly precisely where it started and the road that it took to get there. It is tempting to conclude that irreversibility in general is the consequence of chaotic dynamics, but that remains to be shown.

In situations where the dynamics effectively lead to a loss of information, whether as a result of chaos, irreversibility, or openness, one must use other methods to learn about the past of the system. In this context, openness can have positive as well as negative implications for retrodiction. In our previous example, the expansion of the sphere of influence of the original sound can make it easier rather than harder to recover it after a time has elapsed. Suppose, for example, that the sounds, upon leaving the room in which they were uttered, reach the ears of another person while they are still loud enough to be audible. In this case, a new trace of them will have been made which will be recoverable for a much longer time than the original one was. Even if there is no human listener to hear the words, one can imagine nonliving objects that might be affected by the sound in such a way as to record the words for future detection. An example of this was the suggestion a few years ago that certain ceramic utensils might have encoded fragments of conversations that took place centuries ago while the utensils were being made. If the universe were dense with objects that can act as recording instruments, then as time passed the chances would become increasingly great that any event would eventually influence one of these instruments, and so be recorded. There is an analogy for prediction to the problem that an ever wider set of objects must be considered for retrodiction. When we try to predict what will happen in a specific place, then as we go further into the future an ever greater number of objects in the present may influence this future, and so its prediction requires a knowledge of more and more about the present.

VIII. RECORDS

So far as we know, the influence of most events does not eventually happen upon some medium that will make a stable record of the event. It may be, however, that we are unaware of some of the ways in which such traces can be formed and eventually read. Scientists often find new ways of making inferences about the past based on novel forms of recording. A recent example is the suggestion that there may be evidence of the supernovas that have taken place in our galaxy over an extended period of time recorded in the abundances of certain rare isotopes that have been produced on earth by the strong flux of neutrinos that these supernovae emit.

What are the general characteristics of systems that can produce records of events that stimulate them? By a *record,* we shall mean some configuration of a physical system, usually fairly small, from which, by using generally accepted ideas such as physical laws, we can infer with high probability that a specific event occurred at particular time and place in the past. Records are easy to exemplify with human artifacts. An example would be a videotape showing the Mt. St. Helens volcano erupting at a time identified by a specific configuration of planets in the background sky. Records can also exist without involving human artifacts, although those are often less time specific. An example is the footprint of some animal on a mud flat, which has become preserved through the drying out of the mud. This record indicates that the animal has passed, but not when. But one could, for example, imagine some analysis that would correlate the age of the footprint with some of its current physical characteristics, which would allow for the dating of the animal's passage as well. Something similar to this can actually be done using radioactive dating.

What is it about physical systems that allows them to be used as records?[12] At least two characteristics are required.

1. The system should have various states—the "recording states"—that can be produced by different stimuli. In other words, it should not be that all stimuli produce the same record. Of course, ideally, the system should have many states that are variously produced

---

[12] The modern scientific consideration of the nature of records was initiated by Hans Reichenbach in *The Direction of Time* (Berkeley. California UP, 1956) More recent discussions can be found in Paul Horwich, *Asymmetries in Time* (Cambridge: MIT, 1988), and references therein.

depending on interesting aspects of the stimuli. This requirement can be satisfied to varying degrees. Even for humanly designed recorders, such as particle detectors, it is possible for several different particles to trigger the same response. For a naturally occurring system to be useful as a detector, one can require only that the ambiguity among stimuli is relatively small.

2. The recording states must remain "stable" for a relatively long time so that the record persists and can be read. Stability is required against change to another state that would be interpreted as the result of a different stimulus. There are a number of ways to assure stability. The need for stability seems to require that any physical system that can act as a recorder should have a discrete set of recording states with some strong barrier against spontaneous changes between states. The barrier is frequently a requirement of a large amount of energy to induce a transition between the states. Otherwise, small perturbations from the environment will tend to cause the recording system to make transitions between the recording states, thus destroying the record.

Normally, recording devices designed by humans have a metastable ready state (an unexposed photographic film, for example), so that the interaction with an external system has a high probability of changing it into another state. The recording states (developed film, for example) are more stable, so that once in such a state the recording system does not easily emerge from it. The ready state makes it possible to record weak stimuli, even when the barrier between recording states is a high-energy barrier. But it is not necessary to have a ready state. For example, we use the state of the surface of the moon to estimate the rate of meteor impacts, even though it takes no less energy to crater the unmarked surface than it does subsequently to put new craters on the already marked surface.

Suppose such a recording system interacts with another system external to it, and makes a transition into one of its stable states. Then some time later, inspection of the recording system will allow for the inference that such an interaction has taken place in the past, although it will usually not specify when the interaction occurred. Of course, some theory is involved in this inference, even in a case as simple as determining which animal is responsible for an ancient footprint on a beach. Generally, the more complex the recording system, that is, the more distinct recording states it has, the better the chance of inferring the precise nature of the stimulus that produced the record. Alternatively, it may be necessary to use input from several recording systems in order to narrow down the possible stimuli. It would be very interesting to describe all of the physical systems that can act as recording systems. Such a description, together with an idea of the rate at which external systems stimulate the recording systems, could lead to an estimate of the total rate at which events are being recorded.

In any event, these are patently in the nature of very preliminary remarks. A rigorous, comprehensive, and foundational investigation of the nature of records, of the conditions under which records become possible, and of the important question of why these conditions should obtain only in one time direction and not in the other has yet (so far as we know) to be undertaken.

IX. RETRODICTION IN QUANTUM MECHANICS

The process of retrodiction is importantly different in quantum mechanics than in the macroscopic cases that we have been discussing so far. One reason for this is the fundamentally probabilistic effect of carrying out a *measurement* on a quantum mechanical system. The crucial point is that laws governing the forward-in-time evolution of the quantum state of the world through a measurement event are (on the standard formulation) *probabilistic* laws, and so (in contrast with the case of Newtonian mechanics) those laws of forward-in-time evolution through a measurement will entail *no laws at all* (probabilistic or otherwise) for the *backward-in-time* evolution of the quantum state of the world through such an event.

Suppose, for example, that the position of a certain particle is measured at a certain time. The outcome of that measurement *cannot* (in the absence of other information) be used to infer the position of that particle, or even its probable position, or even to infer probabilities that any of its observable properties had any particular values, at times (no matter how close) prior to the measurement.

In general, according to quantum mechanics, we can only make statements about the situation between the times that measurements have been made, and even these statements are different than they are for predictions. Furthermore, deciding when something has happened is much more problematic in quantum mechanics than it is in Newtonian mechanics. For an isolated system with a definite energy, evolving according to the Schrödinger equation, quantum mechanics states that nothing actually happens. The average value of any physical quantity is independent of time. This is quite different than the corresponding situation in Newtonian mechanics, where, for example, the position of a particle can change even though its energy is constant. Of course, only certain quantities have definite physical values in the quantum mechanical situation, those which can be measured simultaneously with the energy. For these quantities, once their value is known at some time, the situation is about the same as in Newtonian mechanics. But for other quantities, such as the position of an electron in an atom, in a stationary state there is no evolution similar to what happens in Newtonian mechanics.

If we want to discuss retrodiction in a stationary state for quantities that cannot simultaneously be measured with the energy, we might imagine that we could make use of the language of counterfactuals. That is, we can ask what would have been the result had a measurement been done at some time in the past, even though such a measurement has not been done in the actual world. For example, we can ask what would have been the result of a position measurement of an electron at some time in the past, knowing that the electron is definitely in a specific stationary state of an atom and has been so since the last measurement was made on it, which we assume was prior to the time for which we are asking the counterfactual question. For this type of question, the answer given by the formalism of quantum mechanics is very similar to the answer it gives for prediction. That is, quantum mechanics states the probabilities that had the measurement been done, it would have resulted in various values of the position for the electron.

This similarity in formalism, however, hides very important conceptual differences between the predictive and retrodictive questions. If the position measurement has not been performed in the past, then there is no further possibility of doing it, so that the counterfactual retrodiction is entirely sterile, while for the prediction of future positions the possibility exists of actually carrying out the measurement and so testing the prediction. Moreover, if the last measurement made on the electron had instead been a position measurement, then the electron cannot have been in a state of definite energy subsequently. Even if a later measurement of the electron's energy is carried out, then, as stressed earlier, this measurement will not imply that the electron had the energy implied by the measurement beforehand, and therefore the result of the energy measurement cannot be used for retrodiction in the simple way that it can be used for prediction.

In some sense, if a quantity has not been measured in the past, then according to quantum mechanics it has no value at all. One might go so far as to say that according to quantum mechanics, at least in its standard version, the past consists of a discontinuous series of events when measurements have taken place, with nothing happening in the intervals between these measurements. A good analogy to the quantum mechanical picture of the past is a room that is filled with objects but is usually darkened. From time to time, a light turns on briefly illuminating the objects, but then goes out, returning them to the darkness.

This suggests that, in quantum mechanics, retrodictions should be of a different type than we have considered thus far. It appears that in order to make valid retrodictions in quantum mechanics, it is necessary to make use of actual measurements at two different times. Suppose that for a particle, one quantity, such as position, has been measured at time $-T$. If we know the results of that measurement, we can then infer the value of the time-evolved quantity corresponding to position at all subsequent times until another measurement has been made. Suppose, however, that we do not know the result of the position measurement, but only that it has been made. All that we know is that the particle had a definite value of position directly after the measurement. We can ask what further measurements to be done at $t = 0$ will allow us to make precise statements about the particle at times between $-T$ and 0. The answer is that, if we measure one specific quantity, which is the position variable evolved from $-T$ to $t = 0$, then from the result of this measurement we can infer the result of the original measurement of position at time $-T$. This is directly analogous to what we can do for single-measurement prediction. We can also, from the same results, infer the value of a sequence of quantities at intermediate times, even if those quantities have not been measured. These quantities are what the position operator evolves into at those times between $-T$ and 0. In general, they are complicated functions of position and momentum, and will not be easy to measure. These quantities, however, and their analogues for other pairs of measurements, appear to be the only ones for which we can have precise past knowledge "counterfactually," that is, knowledge that does not depend on some measurement having been made at the specific past time that the knowledge is sought.

In some sense, quantum mechanics implies that the past does not exist, except as a result of whatever measurements have been carried out, and which we can do now and in the future. Since we no longer have any control over what measurements have been done in the past, then the only thing that we can still influence is what measurements we will do now and in the future. A proper quantum mechanical reconstruction of the past would involve the determination of as much as we can on the basis of the sequence of measurements that have actually been carried out. If we make a suitable present measurement, we can use the double-measurement retrodiction which has just described to infer something definite about the system between the time of the last measurement and now. For the other measurements in the sequence which have been made, it will usually not be the case that they are related in the proper way to allow us to use single-measurement retrodiction during the times between them.

All of this assumes that there is some way to tell what measurements have been done in the past on some physical system. In general, it is not possible to do this by examining the physical system itself in the present, at least for microscopic systems. While the state of such a system will be influenced by such past measurements, there is no way to infer uniquely from the state what measurements have been done. Furthermore, the state of a system cannot be determined just by examining the system itself. In most situations, such examination will change the state, and will not allow conclusions about the state prior to the measurement. Usually, in order to determine the state and to decide what measurements have been made, we shall need to examine the measuring instruments that have been used. That is, we shall have to examine the records of past measurements. The notion of a measurement which we are using here is a broad one. It involves any interaction between the system of interest and some other recording system, usually macroscopic, such that by examination of the recording system we can infer some definite characteristic of the system of interest. The record of the measurement is just the condition of the recording system after the interaction has taken place. There seems to be something of an infinite regress here, in that in order to determine the condition of the recording system, we might need to make a measurement on it, etc. For the present discussion, we shall assume that for macroscopic recording systems, this problem does not in fact arise. It is not clear whether and why this should be the case. Indeed, one of the major unanswered questions concerning quantum mechanics is why macroscopic systems enter into definite states, corresponding to definite outcomes, rather than remain as mixtures of several states, whose interpretation is uncertain.

Using this notion of measurement, the prospects seem brighter for reviving within quantum mechanics something like the standard notion of the past. A given physical system will have interacted on many past occasions with recording systems. By examining the present traces of these interaction (the records) we can reconstruct some features of the past conditions of the physical system. This reconstruction will not give us a continuous past, but at least it will allow us to make some definite statements about the system at earlier times. To see how this can be done in at least some instances, consider a situation similar to that first described in a celebrated paper written in 1935 by Einstein, Podolsky, and Rosen.[13] Those authors were trying to bury quantum mechanics rather than to help it, but some of their analysis can nevertheless be used for the latter purpose. They considered two systems that interact leaving some joint property, such as total angular momentum, of the combined system known because of some conservation law that is satisfied during the interaction. Neither system, however, has a definite value for its part of the joint quantity, that is, neither system is in a pure state. The systems then separate and do not interact any further. At a later time, by making a measurement on one of the systems, its value of the joint quantity can be determined, and the value of that quantity inferred for the other system by subtraction from its known value for the total system. If the second system is sufficiently simple, this inference is then sufficient to determine its state, without having made any direct measurement on it. Einstein, Podolsky, and Rosen thought that it was highly suggestive that this state characterized the second system at all times after the two systems stopped interacting. This conclusion is warranted when the measurement on system 1 is a completed measurement, that is, when there is no longer a pure state for the joint system. Einstein, Podolsky, and Rosen were concerned with a somewhat different situation in which the joint system remains in a pure state, and there is still an option of what measurement to carry out on system 1. In that case, the seemingly innocent conclusion that system 2 has been in the pure state after the interaction leads to results that contradict experiment. But that problem does not arise in the case that we are considering, for example, if system 1 is a macroscopic recording instrument. By the mysterious process alluded to earlier through which things happen in quantum mechanics, the situation for this instrument has become definite and so has that for the system with which it once interacted. Therefore, it seems possible by this approach to infer some things about the past of physical systems from measurements done in the present. As an example, if the measuring device (system 1) is one from which we can infer the position of an object that interacts with it, then, if such an interaction has taken place in the past, we can infer the past position of the object by looking at the measuring instrument now. Then by use of the evolution equation for the object, we can infer the value of the evolved position at times subsequent to that interaction, and before the next interaction with a recording system. We cannot infer anything definite about quantities other than the evolved position at these times, however. It would be helpful in our thinking about retrodiction to recast quantum mechanics into a form that refers to such "happenings," rather than to the state vectors of the usual formalism. The beginnings of such an approach have been made by Griffiths and Omnes,[14] but it remains to be carried through in detail.

---

[13] Albert Einstein, Nathan Rosen, and Boris Podolsky, *Physical Review,* XLVII (1935)· 777–80.

X. WHAT WILL EVER BE KNOWN ABOUT THE PAST AND FUTURE?

On the basis of these analyses of what can be observed, predicted, and retrodicted according to quantum mechanics, we can draw some general conclusions about what that theory says will ever be known about the world. Imagine that we somehow were in possession of a list of all the observations that have ever or will ever be done on the world. By observations we mean interactions with recording instruments. From that list of observations, together with the laws of quantum mechanics, we can make inferences about things that happened "in-between" the observations. For example, from the information obtained from the *N*th observation, we can infer what the probability would have been for various results of an additional measurement, if that additional measurement had been made before the $N + 1$st one was. This probability can become a certainty if the observation is of the time evolution of the quantity observed last previously. What we can know in this way is obviously much less complete than the corresponding result in Newtonian mechanics, where we can infer the actual situation at times in between measurements. But, in the words of the poet, it is "all we will know on earth." Whether it is all we "need to know" is another matter, one of response to limits to knowledge, which is a normative rather than a cognitive question.

GERALD FEINBERG
SHAUGHAN LAVINE
DAVID ALBERT

Columbia University

---

[14] R. Griffiths, *Journal of Statistical Physics*, XXXVI (1984), p. 219; R. Omnes, *Annals of Physics*, CCI (1990), p. 354.

COMMENTS AND CRITICISM

IS IT LESS WRONG TO HARM THE VULNERABLE THAN THE SECURE?*

Richard Brook[1] has recently offered an account of the distinction between "our common (deontological) morality" and consequentialism. He presents the deontologist as saying this:

> The morally significant feature of not being threatened at a moment demands some special protection, a protection embodied in the right not to be harmed (203).

This morally significant feature, which is possessed by some moral patients but not others, grounds a moral duty. It is more wrong, Brook argues, to harm someone who is already threatened than someone who is not so threatened. To be threatened is to be such that (roughly) "no causal chain has been initiated which, if uninterrupted, will imminently lead to its harm" (202). Brook offers this both as something the deontologist would say, and something he himself believes. The consequentialist, on the other hand, will accept no such distinction, but will prefer to channel harm in whatever direction will produce the best consequences. Brook motivates his account with an example:

> You are at the zoo with two children who are making a scene. Becoming angry, you toss them into the lion's den. Horrified, you come to your senses and notice that they can only be saved if you toss a third child (who just toddled along) into the back of the den. The beast would be distracted and you could quickly leap in and save the first two. What should you do? . . . it seems clearly wrong to kill the third child . as wrong, in fact, as in any of the standard cases where you might harm one person to aid others (197–8).

I do not wish to take a stand on whether Brook is correct in thinking that it is wrong to kill the third child, but I agree that it is probably wrong *if* the standard cases of killing one to save others are wrong. For present purposes, what matters is that the deontologist will say that it is wrong, and I am interested in Brook's account as a principle from which to derive deontological intuitions. His account gets the correct answer from the deontological point of view. When the parent considers throwing in the third child, he is not threat-

---

* I am indebted to Daniel Howard-Snyder for help with this paper

[1] In "Agency and Morality," this JOURNAL, LXXXVIII, 4 (April 1991): 190–212

0022-362X/92/8912/643–7

ened, whereas the first two are. If those who are unthreatened are worthy of some special protection not granted to the threatened, then it is wrong to throw in the third child, even if doing so is necessary to save his siblings.

I

If Brook is right, there should be no moral dilemmas to which his account gives one answer, while deontologists confidently and unanimously give the opposite answer. In this section, I argue that there are a number of such dilemmas, and that a different account (which I call the positive/negative account) explains the data better than Brook's.

Consider the following two cases:

> I. Jack has a kind of cancer which makes it necessary to remove his arm. In other words, he is *threatened* with losing his arm. Now suppose (to return to a familiar example) a doctor working in the hospital realizes that Jack's organs could be used to save the lives of five healthy patients. Is it permissible for her to chop Jack up and take his organs? Is it more permissible to chop him up than to chop up a completely healthy patient, Jill, who has come to the hospital for a yearly check-up?
>
> II. Alice and Bob are going on separate trips into the desert. The terrain they will cover is not particularly hostile, except that it is extremely hot and there is no water available. If a traveler has two bottles of water, there is a very high probability he will survive the trip. If he has one bottle, there is a fifty-fifty chance he will survive; and if he has no water, he will certainly die. Suppose that unbeknown to Alice, one of her bottles has a leak and all the water has run out. Now suppose that you are walking through the desert dying of thirst. You find Alice lying asleep and consider drinking the water in her bottle. Realizing that this is all the water she has, you feel uncomfortable about drinking it and walk off to wrestle with your conscience. A little while later you come across Bob, and notice that he has two bottles. You are determined to steal one of the three bottles, but you debate with yourself which action would be least wrong. Would it be worse to drink Alice's water or to drink Bob's?

No deontologist worthy of the name would say that it is more acceptable, in the first case, to chop up Jack than to chop up Jill. Nor would he say, with respect to the second case, that it is more acceptable for you to drink Alice's water than drink Bob's. That is, in these two cases, a deontologist would not agree that it was more acceptable to harm the person who is already threatened than the person who is not.

But Brook's account leads us to believe that this is exactly what the deontologist would say. For example, in the second case, it seems as if Brook should say that it is preferable to drink Alice's water, because if you were to drink Bob's you would be creating a threat of death for someone who is otherwise unthreatened; whereas if you drink Alice's water you are simply harming someone who is already threatened.[2]

Brook has missed the point of his own examples. It seemed, in his lion's den example, that it *was* less wrong to allow the children I just threw into the lion's den to die than to kill another child to save them. In my example, however, it seemed that it *was not* less wrong to kill the man whose arm was about to be amputated than to kill the unthreatened bystander. (And this is so in spite of the fact that the man whose arm is about to be amputated is already threatened with some harm.)

Killing Jack to get his organs has something in common with allowing the two children to be devoured by the lion: each is a case of choosing that a harm come to someone who is threatened rather than someone who is not threatened. If Brook were correct, we should say the same about the hospital case as we say about the lion case: that it is better for us to choose that the threatened person or people suffer. But we do not. Our intuitions about the cases differ.

Killing Jack to get his organs differs from allowing the two children to be devoured by the lion in that the first is a case of a doing, and the second is a mere allowing. Is it not reasonable to suppose that that is what explains the difference in our intuitions about the cases?

It will no doubt be objected that there are other differences between the cases that might explain the differences in our reactions. In Brook's case, the agent was *forced to choose* between letting the two children be killed by the lion and throwing in the third child. In the case I described, the doctor is not forced to choose between chopping up Jack and chopping up Jill. There is a third choice. She can chop up neither. This difference is itself partly a function of the fact that Brook's example involves a choice between doing and allowing (there is no third alternative to doing P or not doing P), whereas mine involves a choice between doing one thing and doing another.

---

[2] Being at *risk* of imminent death does not quite fit Brook's definition of being threatened, since it is not clear that there is a causal chain which, if uninterrupted, will lead to one's death in this case. But I think it would be fair to expand Brook's point to treat this as a threat. We can see this by adapting his original example slightly. Suppose it is less than certain that the lion will eat the children. Does it follow that they are unthreatened, and if unthreatened no less worthy of special protection than the other child? It is extremely implausible to say that it is permissible to throw in the youngest child in the case in which it is less than certain that the other two will be eaten, whereas it is impermissible in the case in which it is certain that they will be eaten

The doctor in my example *is* forced, however, to choose one of the following three options: chopping up Jack, chopping up Jill, and failing to save the five who are suffering from organ failure. If she ought not to chop up Jack or Jill, then Brook is still in trouble. His theory should not have the result that it is worse that harm happen to Jack, someone who is threatened, than that a comparable harm happen to five other threatened people.

There is one more difference between chopping up Jack and letting the children be eaten by the lion. Jack is not *threatened with death*, whereas the children are. Maybe Brook's point should be that those who are unthreatened in a certain *respect* deserve special protection from threats or harms of a comparable kind.

But suppose that instead of wanting to chop up Jack, the doctor wants to amputate his healthy arm to use it in her experiments. The threat of losing one's left arm is comparable to the threat of losing one's right arm. So, if some considerable evil can be averted by amputating Jack's arm, Brook should say that the doctor may do so. But the deontologist would not agree, I conjecture. Or consider the choice between drinking Alice's water and drinking Bob's. Raising the probability of someone's dying of thirst from .5 to 1 is comparable to raising it from 0 to .5. If Brook is right, it should be slightly more acceptable to increase the chance of Alice's dying than to increase the chance that Bob will die. But it does not seem better to drink Alice's water than to drink Bob's.

The only difference between Brook's example and mine which explains the difference in our intuitions about them is that his is a case of allowing harm and mine are cases of doing harm. It appears, therefore, that the principle that it is worse to kill than to let die, or to harm than to allow to be harmed (call this the positive/negative account), explains the intuitive data considerably better than Brook's account.

II

If I am right, and the positive/negative account really does get at the essence of deontology better than the "threatened/nonthreatened" account that Brook offers, then the positive/negative account should get the deontologically correct answer in the lion's den case.

Unfortunately, it does not. The positive/negative account will say that it is worse to kill than to let die. Assuming that it is worse to kill twice than to kill once, it will tell the parent to throw his youngest child into the lion's den, because in doing so he will be killing one child, but thereby preventing two killings of which he would otherwise be the agent. He is faced with a choice not between killing one and letting two die, but between killing one and killing two. Even if we deny the plausible assumption that killing one is less wrong than killing two, it remains at least permissible to throw the youngest child in. Killing one cannot be more wrong than killing two.

Brook makes this point himself, and offers the advocate of the positive/negative account this amendation: "*at a given moment* one is more responsible not to cause harm than to allow harm to occur" (194). The important difference here is the emphasis on 'at a given moment'. This new version does get the correct result in the lion's den case. The agent has to choose between now killing his youngest child, and it being true that he killed the older two in the past. Another way to describe his choice is this: *now causing* the death of his youngest, as opposed to *now allowing* his other children to die. This account says that it would be wrong for him now to cause the death of anyone. So, it would be wrong to throw in the little boy.

Brook prefers his own account:

> It is the moral significance of that [unthreatened] status that accounts for the fact that the agent's duty must be temporally indexed. Otherwise, we would wonder why a person especially concerned with her own agency would not wish to minimize her own wrongdoing overall (i.e., in a temporally unrestricted way) (210).

His point is that either the temporally indexed version of the positive/negative account is correct but supervenient on his account, or it is incorrect. He believes that that account, offered by itself and not as derivative from some more fundamental view, would be ad hoc and hence implausible.

It is hard to know what weight to give such a consideration. It is true that as we construct theories we are concerned that our principles not be ad hoc or overly complicated. But we are more concerned that our theories account for important data. The temporally indexed version of the positive/negative account does that, and does so considerably better than Brook's account.

The temporally indexed aspect of the positive/negative distinction does seem a bit ad hoc. It does not seem very plausible all by itself as a moral principle, aside from the fact that it gets the correct answers in a number of cases. But then Brook's own account does not look very plausible all by itself. It is not very plausible that being unthreatened should give someone a right not to be threatened or harmed. Would it not be rather odd if morality told us that the weak and vulnerable were more deserving of abuse than the strong and secure?

FRANCES HOWARD-SNYDER

Purdue University

BOOK REVIEWS

Constructibility and Mathematical Existence. CHARLES S. CHIHARA. New York: Oxford University Press, 1990. xv, 282 p. Cloth $39.95.

Charles S. Chihara concurs with most contemporary philosophers of mathematics that Platonism about mathematical objects faces serious epistemic problems. From here his disagreement with them is deep and far reaching—even to the extent developing in "Part I, The Constructibility Theory," "another kind of mathematics" (23). He supports this seemingly radical course in "Part II, Philosophical Developments," where while criticizing contemporary views he argues that his own avoids the problems upon which they founder. This scrutiny covers various realisms (W. V. Quine, Kurt Gödel, J. Burgess, and Penelope Maddy), structuralism (S. Shapiro and myself), Hartry Field's nominalism, and Philip Kitcher's ideal constructor. It also provides a valuable survey of much current philosophy of mathematics. I shall forgo summarizing this or evaluating his criticisms so that I can concentrate upon his positive views.

Chihara's "alternative mathematics" replaces set-existence assertions by claims that open sentences are constructible. For this to work, ordinary mathematics must be formulated within the simple theory of types. This allows us to map 'Individual x belongs to a type 1 class A' onto a Chihara sentence 'x satisfies F', where F is an open sentence token with extension A. We map 'A belongs to the type 2 class X' onto 'Token F satisfies the open sentence token K', where the extension of K is X, and so on. Replacing classes of higher type by open sentences raises problems with identities, since classes are extensional. Chihara handles this by restricting the domain of open sentences to those for which interchange of extensionally equivalent arguments holds. (This also deflects problems with quantifying into opaque contexts.) He works out the formal details in two chapters —essentially transferring simple type theory into the new setting.

Satisfaction is a primitive here; for obvious reasons, he does not define it in terms of classes. He admits that this means that his ideology is not fully nominalist. His ontology remains so, however, since his constructibility claims concern sentence tokens, not their types. Although he bans possible worlds as abstract entities, he uses

0022-362X/92/8912/648–51

the heuristic of thinking of constructibility claims as saying that the open sentences in question are tokened in some possible world. Furthermore, the people doing the tokening in those worlds might use a language that we could not grasp or understand. This is supposed to allow constructibility theory to produce correlates of uncountable infinities.

A chapter shows how the Frege-Russell cardinal numbers can be recovered straightforwardly within constructibility theory. Its successor contains a novel reconstruction of the theory of measurable quantities, using ordinary quantifiers ranging over one-dimensional objects having length. Chihara does not commit himself to these "strange entities" (102) but rather offers his theory as an idealization. Aside from its intrinsic interest, Chihara seems to have included this theory in order to show that, contrary to the currently influential Quine-Putnam indispensability argument, we need not refer to mathematical entities in doing science. Now, I grant that we can substitute idealized entities for real valued functions. But I wonder whether mathematics will intrude again in applying this theory. Henry Kyburg's *Theory and Measurement,* which treats quantities along lines similar to Chihara's, gives a detailed analysis of the inferences required to apply them in measurement. His account still requires a statistical error theory based upon real analysis.

Of course, an error theory of this sort can be developed within Chihara's system or an extension of it. Thus, he can still reject indispensability arguments, as he also does, on the ground that in doing science we can avoid committing ourselves to abstract mathematical objects by interpreting mathematics in terms of constructibility theory.

This raises the question of the relationship between Chihara's system and ordinary mathematics. Chihara tries repeatedly to clarify this, but I remain puzzled. In calling his an alternative mathematics he refers approvingly to A. Heyting (23), but he is not out to reform mathematical practice. To accommodate it he has even relaxed the less controversial version of constructibility presented in his previous book, *Ontology and the Vicious Circle Principle.* Furthermore, he argues against Field that standard mathematics "does contain much truth" (173), and states that it is consistent for him to hold that it is at least approximately true if properly construed. In illustrating this thesis, he claims that "what is essentially Lebesque's existence theorem" can be proven within constructibility theory, thereby showing it to be "correct and sound" (258). This fits poorly with his denial that he is analyzing the meaning of mathematical statements. For by itself Chihara's mapping of standard mathematics into constructibility theory shows only that the former is consistent relative to the latter. The mapping must preserve more for us to conclude that it establishes that mathematics is also (approximately) true. Chihara does not tell us what this is. Nor do I see how he can without facing a number of familiar difficulties confronting reductionist programs—ones that he emphasizes do not touch his own. These include a version of Paul Benacerraf's multiple reduction problem, since type theory still leaves us many options for defining, say, the reals.

In criticizing Burgess's realism, Chihara says: "I want to treat the question of the existence of mathematical entities as a scientific question . . . Where is the evidence that supposedly supports belief in these entities? Here, I mean evidence of the sort that would convince the physicist or biologist" (189). This prompts me to wonder about the evidence we might seek for constructibility claims. Take the case of the axiom of choice. Its correlate in Chihara's system denies that an open sentence is constructible for which no open sentence serving as its "choice function" is constructible. Although this is not one of his axioms, there is good reason for including it, for the usual axiom is now standard mathematics. The considerations mathematicians have raised historically for and against it, however, are irrelevant to Chihara's version. He has already disavowed the constructivist objection that specifying some choice functions might require means that we could grasp. Nor can mathematical fruitfulness count in its favor, since that has nothing to do with the sorts of open sentences that might be tokened. Rather, to seek evidence bearing upon the Chihara principle we would need to turn to biology, anthropology, and linguistics in order to determine what sorts of languages possible humans might have. Furthermore, we should be wary of idealizations of ourselves; Chihara complains that Kitcher's ideal constructor "appears more godlike than human" (243).

We would find, I suspect, that the open sentences possible humans might token are limited by the complexity of their extensions. This restriction might make Chihara's choice principle hold trivially, just as restrictions on the domain of intuitionist mathematics yield trivial versions of the axiom. But if so, constructibility theory would also be limited, and probably could not provide proxies for infinitely many complicated sets of integers. In standard mathematics, we sometimes leap over barriers due to complexity by appealing to pure existence theorems. Thus we know, for instance, that uncountably many sets of integers are not definable in the notation of ZF set theory. If the system is consistent, then some of these are extensions of "truth predicates" for ZF. But we can easily add such a predicate, by adding for each numeral an axiom of the form '*Tn* if and only if *S*', where *n* Gödel numbers *S* under some Gödel numbering we define in ZF in advance. One can do something analogous in constructibility theory. But this only postpones the problem. We cannot define all the Gödel numberings theoretically definable in ZF; there are infinitely many and some are too complicated. Despite this, our mathematics tells us that they exist. Can the constructibility theorist say something similar? Until we know more about how complexity limits constructibility, we just cannot say.

Perhaps I have misunderstood Chihara's theory, for I have taken constructibility theory as concerned with what it is logically possible for possible humans to do. By contrast, it seems to me that in approaching mathematics in constructive terms the properties of the operations generating possible constructions are of essence, not those of the agents using them. This makes constructibility a purely mathematical matter rather than biolinguistic one.

However these and other questions concerning Chihara's theory are resolved, readers will find much of interest in this book.

MICHAEL D. RESNIK

University of North Carolina/Chapel Hill

INDEX TO VOLUME LXXXIX

INDEX OF SUBJECTS DISCUSSED IN THE JOURNAL

Abstract Objects

Hale, Bob: Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. 89 (3): 111–35 '92

Wright, Crispin: Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. 89 (3): 111–35 '92

Action

Driver, Julia: Caesar's Wife: On the Moral Significance of Appearing Good. 89 (7): 331–43 '92

Agency, Social

Howard-Snyder, Frances: Is It Less Wrong to Harm the Vulnerable than the Secure? 89 (12): 643–7 '92

Bayesian Conditionalization

Dorling, Jon: Bayesian Conditionalization Resolves Positivist/Realist Disputes. 89 (7): 362–82 '92

Belief

Hansson, Sven Ove: In Defense of the Ramsey Test. 89 (10): 522–40 '92

Sartwell, Crispin: Why Knowledge is Merely True Belief. 89 (4): 167–80 '92

Schiffer, Stephen: Belief Ascription. 89 (10): 499–521 '92

Brook, Richard

Howard-Snyder, Frances: Is It Less Wrong to Harm the Vulnerable than the Secure? 89 (12) 643–7 '92

Conditionals

Hansson, Sven Ove· In Defense of the Ramsey Test. 89 (10): 522–40 '92

Davidson, Donald

Bovens, Luc: Sour Grapes and Character Planning. 89 (2): 57–78 '92

Washington, Corey: The Identity Theory of Quotation. 89 (11): 582–605 '92

Decision Theory

Kusser, Anna The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn. 89 (1): 10–29 '92

Spohn, Wolfgang: The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn. 89 (1): 10–29 '92

Deontology

Howard-Snyder, Frances: Is It Less Wrong to Harm the Vulnerable than the Secure? 89 (12): 643–7 '92

Distinction, Standard and Nonstandard

Hintikka, Jaakko: The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. 89 (6): 290–315 '92

Sandu, Gabriel· The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. 89 (6): 290–315 '92

Character Planning

Bovens, Luc: Sour Grapes and Character Planning. 89 (2). 57–78 '92

Community

Meyer, Michael J.: Rights between Friends. 89 (9). 467–83 '92

Elster, Jon

Bovens, Luc. Sour Grapes and Character Planning. 89 (2): 57–78 '92

Envy

Ben-Ze'ev, Aaron: Envy and Inequality. 89 (11) 551–81 '92

Epistemology
- Sartwell, Crispin: Why Knowledge Is Merely True Belief 89 (4): 167–80 '92
- Van Cleve, James: Semantic Supervenience and Referential Indeterminacy. 89 (7): 344–61 '92

Field, Hartry
- Hale, Bob: Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. 89 (3). 111–35 '92
- Wright, Crispin: Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. 89 (3): 111–35 '92

Frege, Gottlob
- Hintikka, Jaakko: The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. 89 (6): 290–315 '92
- Sandu, Gabriel. The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction By Jaakko Hintikka and Gabriel Sandu. 89 (6): 290–315 '92

Friendship
- Meyer, Michael J.: Rights between Friends. 89 (9): 467–83 '92

Gärdenfors, Peter
- Hansson, Sven Ove: In Defense of the Ramsey Test. 89 (10) 522–40 '92

Gauthier, David
- Yi, Byeong-Uk: Rationality and the Prisoner's Dilemma in David Gauthier's *Morals by Agreement*. 89 (9): 484–95 '92

Harm
- Howard-Snyder, Frances: Is It Less Wrong to Harm the Vulnerable than the Secure? 89 (12): 643–7 '92

Indeterminacy
- Ross, James: Immaterial Aspects of Thought. 89 (3): 136–50 '92

Indispensability Arguments
- Maddy, Penelope: Indispensability and Practice. 89 (6): 275–89 '92

Inequality
- Ben-Ze'ev, Aaron: Envy and Inequality. 89 (11)· 551–81 '92

Knowledge
- Sartwell, Crispin: Why Knowledge Is Merely True Belief. 89 (4): 167–80 '92

Loopholes
- Pogge, Thomas W.: Loopholes in Moralities. 89 (2): 79–98 '92

Metaphysics
- Price, Huw: Metaphysical Pluralism. 89 (8): 387–409 '92

Miller, Richard B.
- Hellman, Geoffrey: Supervenience/Determination a Two-way Street? Yes, But One of the Ways Is the *Wrong Way!* 89 (1): 42–7 '92

Moral Philosophy
- Ben-Ze'ev, Aaron: Envy and Inequality. 89 (11): 551–81 '92
- Bittner, Rüdiger: Is It Reasonable to Regret Things One Did? 89 (5): 262–73 '92
- Driver, Julia: Caesar's Wife: On the Moral Significance of Appearing Good. 89 (7): 331–43 '92
- Pogge, Thomas W.: Loopholes in Moralities. 89 (2)· 79–98 '92

Naturalism
- Quine, W. V.. Structure and Nature. 89 (1): 5–9 '92

New Books.
- 89 (4): 210–6; (10): 541–9

New Books: Anthologies.
89 (4): 216–9; (8). 443–4; (10): 549–50

New Books: Paperbacks.
89 (4)· 219–20; (8): 441

New Books: Translations.
89 (4): 220–2, (8): 441–2

Nominalism
Hale, Bob: Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. 89 (3): 111–35 '92
Wright, Crispin: Nominalism and the Contingency of Abstract Objects By Bob Hale and Crispin Wright. 89 (3): 111–35 '92

Notes and News.
89 (4)· 209–10; (5). 274; (6): 330; (9): 496–8; (11): 606

Ontology
Sidelle, Alan: Rigidity, Ontology, and Semantic Structure. 89 (8): 410–30 '92

Paradox, Liar
Gaifman, Haim: Pointers to Truth. 89 (5) 223–61 '92

Philosophy of Language
Dudman, V. H.: A Popular Presumption Refuted. 89 (8): 431–2 '92
Schiffer, Stephen· Belief Ascription. 89 (10): 499–521 '92
Washington, Corey: The Identity Theory of Quotation. 89 (11)· 582–605 '92

Philosophy of Logic
Hintikka, Jaakko The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. 89 (6). 290–315 '92
Sandu, Gabriel. The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. 89 (6): 290–315 '92

Philosophy of Mathematics
Gaifman, Haim. Pointers to Truth. 89 (5): 223–61 '92
Hale, Bob: Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. 89 (3): 111–35 '92
Hintikka, Jaakko: The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu 89 (6): 290–315 '92
Maddy, Penelope· Indispensability and Practice. 89 (6). 275–89 '92
Quine, W V.: Structure and Nature. 89 (1): 5–9 '92
Sandu, Gabriel: The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. 89 (6): 290–315 '92
Wright, Crispin: Nominalism and the Contingency of Abstract Objects By Bob Hale and Crispin Wright. 89 (3): 111–35 '92

Philosophy of Science
Albert, David Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12). 607–42 '92
Dorling, Jon: Bayesian Conditionalization Resolves Positivist/Realist Disputes 89 (7): 362–82 '92
Feinberg, Gerald: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Lavine, Shaughan: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92

Physicalism
Hellman, Geoffrey: Supervenience/Determination a Two-way Street? Yes, But One of the Ways Is the *Wrong Way!* 89 (1): 42–7 '92

Pluralism, Metaphysical
Price, Huw: Metaphysical Pluralism. 89 (8) 387–409 '92

Positivism
Dorling, Jon. Bayesian Conditionalization Resolves Positivist/Realist Disputes. 89 (7): 362–82 '92

Practical Deliberation
Kusser, Anna: The Utility of Pleasure is a Pain for Decision Theory. By

Anna Kusser and Wolfgang Spohn. 89 (1): 10–29 '92
Spohn, Wolfgang: The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn. 89 (1): 10–29 '92

Practice, Mathematical

Maddy, Penelope· Indispensability and Practice. 89 (6): 275–89 '92

Prediction

Albert, David: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Feinberg, Gerald: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Lavine, Shaughan: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92

Preferences

Bovens, Luc: Sour Grapes and Character Planning. 89 (2) 57–78 '92

Prisoner's Dilemma

Yi, Byeong-Uk· Rationality and the Prisoner's Dilemma in David Gauthier's *Morals by Agreement.* 89 (9): 484–95 '92

Propositional Attitudes

Seymour, Michel: A Sentential Theory of Propositional Attitudes. 89 (4) 181–201 '92

Quantum Mechanics

Albert, David: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Feinberg, Gerald: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Lavine, Shaughan: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92

Quotation

Washington, Corey: The Identity Theory of Quotation. 89 (11): 582–605 '92

Rational Choice

Schick, Frederic: Allowing for Understandings. 89 (1): 30–41 '92
Schmidtz, David: Rationality within Reason 89 (9): 445–66 '92
Yi, Byeong-Uk: Rationality and the Prisoner's Dilemma in David Gauthier's *Morals by Agreement.* 89 (9): 484–95 '92

Rationality

Schmidtz, David· Rationality within Reason. 89 (9)· 445–66 '92

Ramsey Test

Hansson, Sven Ove: In Defense of the Ramsey Test. 89 (10): 522–40 '92

Realism

Dorling, Jon: Bayesian Conditionalization Resolves Positivist/Realist Disputes 89 (7). 362–82 '92

Referential Indeterminacy

Van Cleve, James: Semantic Supervenience and Referential Indeterminacy 89 (7): 344–61 '92

Regret

Bittner, Rüdiger: Is It Reasonable to Regret Things One Did? 89 (5): 262–73 '92

Retrodiction

Albert, David: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Feinberg, Gerald: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92
Lavine, Shaughan: Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92

Rights

Meyer, Michael J.: Rights between Friends. 89 (9): 467–83 '92

Rigid Designation

Sidelle, Alan: Rigidity, Ontology, and Semantic Structure. 89 (8): 410–30 '92

Semantics

Dudman, V. H.: A Popular Presumption Refuted. 89 (8): 431–2 '92
Gaifman, Haim: Pointers to Truth. 89 (5): 223–61 '92
Schiffer, Stephen: Belief Ascription. 89 (10): 499–521 '92
Seymour, Michel: A Sentential Theory of Propositional Attitudes. 89 (4): 181–201 '92
Sidelle, Alan: Rigidity, Ontology, and Semantic Structure. 89 (8): 410–30 '92
Van Cleve, James: Semantic Supervenience and Referential Indeterminacy. 89 (7): 344–61 '92
Washington, Corey: The Identity Theory of Quotation. 89 (11): 582–605 '92

Sentential Theory

Seymour, Michel: A Sentential Theory of Propositional Attitudes. 89 (4): 181–201 '92

Structuralism

Quine, W. V.: Structure and Nature. 89 (1): 5–9 '92

Supervenience

Hellman, Geoffrey: Supervenience/Determination a Two-way Street? Yes, But One of the Ways Is the *Wrong Way!* 89 (1): 42–7 '92
Van Cleve, James: Semantic Supervenience and Referential Indeterminacy. 89 (7): 344–61 '92

Thought

Ross, James: Immaterial Aspects of Thought 89 (3): 136–50 '92

Truth

Gaifman, Haim: Pointers to Truth. 89 (5): 223–61 '92

Understanding

Schick, Frederic: Allowing for Understandings. 89 (1): 30–41 '92

Utility

Kusser, Anna: The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn. 89 (1): 10–29 '92
Schick, Frederic: Allowing for Understandngs. 89 (1): 30–41 '92
Schmidtz, David: Rationality within Reason. 89 (9): 445–66 '92
Spohn, Wolfgang: The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn 89 (1): 10–29 '92

INDEX OF AUTHORS

Including Contributing Authors and Reviewers

Albert, David, Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. 89 (12): 607–42 '92

Baldwin, Thomas, Review of Peter Hylton: ***Russell, Idealism, and the Emergence of Analytic Philosophy.*** 89 (1): 51–5 '92
Ben-Ze'ev, Aaron, Envy and Inequality. 89 (11): 531–81 '92
Berofsky, Bernard, Review of Susan Wolf: ***Freedom within Reason.*** 89 (4): 202–8 '92
Bittner, Rüdiger, Is It Reasonable to Regret Things One Did? 89 (5): 262–73 '92
Bovens, Luc, Sour Grapes and Character Planning. 89 (2): 57–78 '92

Dorling, Jon, Bayesian Conditionalization Resolves Positivist/Realist Disputes. 89 (7): 362–82 '92
Driver, Julia, Caesar's Wife: On the Moral Significance of Appearing Good. 89 (7): 331–43 '92
Dudman, V. H., A Popular Presumption Refuted. 89 (8): 431–2 '92

Feinberg, Gerald, Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. **89** (12): 607–42 '92

Gaifman, Haim, Pointers to Truth. **89** (5): 223–61 '92
Garber, Daniel, Review of Catherine Wilson: *Leibniz's Metaphysics: A Historical and Comparative Study* and Robert C. Sleigh, Jr.: *Leibniz and Arnauld: A Commentary on their Correspondence.* **89** (3): 151–65 '92
Guyer, Paul, Review of Henry E. Allison: *Kant's Theory of Freedom.* **89** (2): 99–110 '92

Hale, Bob, Nominalism and the Contingency of Abstract Objects. By Bob Hale and Crispin Wright. **89** (3): 111–35 '92
Hansson, Sven Ove, In Defense of the Ramsey Test. **89** (10): 522–40 '92
Hellman, Geoffrey, Supervenience/Determination a Two-way Street? Yes, But One of the Ways Is the *Wrong Way!* **89** (1): 42–7 '92
Hintikka, Jaakko, The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu **89** (6): 290–315 '92
Howard-Snyder, Frances, Is It Less Wrong to Harm the Vulnerable than the Secure? **89** (12): 643–7 '92

Kusser, Anna, The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn. **89** (1): 10–29 '92

Lavine, Shaughan
  Review of Penelope Maddy: *Realism in Mathematics.* **89** (6): 321–6 '92
  Knowledge of the Past and Future. By Gerald Feinberg, Shaughan Lavine, and David Albert. **89** (12): 607–42 '92

Maddy, Penelope, Indispensability and Practice. **89** (6): 275–89 '92
McGinn, Colin, Review of Malcolm Budd: *Wittgenstein's Philosophy of Psychology.* **89** (8): 433–6 '92
Meyer, Michael J., Rights between Friends. **89** (9). 467–83 '92

Neuhouser, Frederick, Review of Allen Wood: *Hegel's Ethical Thought.* **89** (6): 316–20 '92

Pogge, Thomas W., Loopholes in Moralities. **89** (2): 79–98 '92
Price, Huw, Metaphysical Pluralism. **89** (8): 387–409 '92

Quine, W. V., Structure and Nature. **89** (1): 5–9 '92

Resnik, Michael D., Review of Charles Chihara: *Constructibility and Mathematical Existence.* **89** (12): 648–51 '92
Ross, James, Immaterial Aspects of Thought. **89** (3): 136–50 '92

Sandu, Gabriel, The Skeleton in Frege's Cupboard: The Standard versus Nonstandard Distinction. By Jaakko Hintikka and Gabriel Sandu. **89** (6): 290–315 '92
Sartwell, Crispin, Why Knowledge Is Merely True Belief. **89** (4): 167–80 '92
Schick, Frederic, Allowing for Understandings. **89** (1): 30–41 '92
Schiffer, Stephen, Belief Ascription. **89** (10): 499–521 '92
Schmidtz, David, Rationality within Reason. **89** (9): 445–66 '92
Seymour, Michel, A Sentential Theory of Propositional Attitudes. **89** (4): 181–201 '92
Sidelle, Alan, Rigidity, Ontology, and Semantic Structure. **89** (8): 410–30 '92
Spohn, Wolfgang, The Utility of Pleasure is a Pain for Decision Theory. By Anna Kusser and Wolfgang Spohn. **89** (1): 10–29 '92
Stroud, Barry, Review of W. V. Quine and Rudolf Carnap: *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work.* **89** (7): 383–6 '92

Van Cleve, James, Semantic Supervenience and Referential Indeterminacy. 89 (7): 344–61 '92

Washington, Corey, The Identity Theory of Quotation. 89 (11): 582–605 '92

Wellman, Carl, Review of Judith Jarvis Thomson: *The Realm of Rights*. 89 (6): 326–9 '92

Williams, Michael, Review of W. V. Quine: *Pursuit of Truth*. 89 (1): 48–51 '92

Woolhouse, R. S., Review of Michael Ayers. *Locke: Volume I, Epistemology; Volume II, Ontology*. 89 (8): 436–40 '92

Wright, Crispin, Nominalism and the Contingency of Abstract Objects By Bob Hale and Crispin Wright 89 (3): 111–35 '92

Yi, Byeong-Uk, Rationality and the Prisoner's Dilemma in David Gauthier's *Morals by Agreement*. 89 (9): 484–95 '92

INDEX OF BOOKS REVIEWED

Listed by Author of Item Reviewed

Allison, Henry E.: *Kant's Theory of Freedom*.—Paul Guyer. 89 (2): 99–110 '92

Ayers, Michael: *Locke: Volume I, Epistemology; Volume II, Ontology*.—R. S Woolhouse. 89 (8): 436–40 '92

Budd, Malcolm: *Wittgenstein's Philosophy of Psychology* —Colin McGinn. 89 (8): 433–6 '92

Carnap, Rudolf: *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*. By W. V. Quine and Rudolf Carnap —Barry Stroud. 89 (7): 383–6 '92

Chihara, Charles: *Constructibility and Mathematical Existence*.—Michael D. Resnik. 89 (12): 648–51 '92

Hylton, Peter: *Russell, Idealism, and the Emergence of Analytic Philosophy*.—Thomas Baldwin. 89 (1): 51–5 '92

Maddy, Penelope: *Realism in Mathematics*.—Shaughan Lavine. 89 (6): 321–6 '92

Quine, W. V.: *Pursuit of Truth*.—Michael Williams. 89 (1): 48–51 '92. *Dear Carnap, Dear Van. The Quine-Carnap Correspondence and Related Work*. By W. V. Quine and Rudolf Carnap.—Barry Stroud. 89 (7): 383–6 '92

Sleigh, Robert C., Jr.: *Leibniz and Arnauld: A Commentary on their Correspondence*.—Daniel Garber. 89 (3): 151–65 '92

Thomson, Judith Jarvis: *The Realm of Rights*.—Carl Wellman. 89 (6): 326–9 '92

Wilson, Catherine: *Leibniz's Metaphysics: A Historical and Comparative Study*.—Daniel Garber 89 (3): 151–65 '92

Wolf, Susan: *Freedom within Reason*.—Bernard Berofsky. 89 (4): 202–8 '92

Wood, Allen: *Hegel's Ethical Thought*. —Frederick Neuhouser. 89 (6): 316–20 '92

Statement of Ownership, Management and Circulation

(Required by 39 U S C 3685)

1A Title of Publication
THE JOURNAL OF PHILOSOPHY

1B PUBLICATION NO
0 0 2 2 3 6 2 X

2 Date of Filing
10/5/92

3 Frequency of Issue
MONTHLY

3A No. of Issues Published Annually
12

3B Annual Subscription Price
$65.00 Institution
$35.00 Individual
~~$20.00 Student~~

4 Complete Mailing Address of Known Office of Publication *(Street, City, County, State and ZIP+4 Code) (Not printers)*
709 Philosophy Hall, Columbia University, New York NY 10027

5 Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*
See #4

6 Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*
The Journal of Philosophy, 709 Philosophy Hall, New York NY 10027

Editor *(Name and Complete Mailing Address)*
Arthur Danto, 709 Philosophy Hall, Columbia University New York, Ny 10027

Managing Editor *(Name and Complete Mailing Address)*
Michael Kelly, 704 Philosophy Hall, Columbia University, New York NY 10027

7 Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Journal of Philosophy	709 Philosophy Hall Columbia University New York, NY 10027

8 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities *(If there are none, so state)*

Full Name	Complete Mailing Address
NONE	

9 For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 424.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) ☒ Has Not Changed During Preceding 12 Months
(2) ☐ Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10 Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A Total No. Copies *(Net Press Run)*	4845	4745
B Paid and/or Requested Circulation 1 Sales through dealers and carriers, street vendors and counter sales	——	——
2 Mail Subscription *(Paid and/or requested)*	3972	3915
C Total Paid and/or Requested Circulation *(Sum of 10B1 and 10B2)*	3972	3915
D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	174	178
E Total Distribution *(Sum of C and D)*	4146	4093
F Copies Not Distributed 1 Office use, left over, unaccounted, spoiled after printing	710	651
2 Return from News Agents	——	——
G TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	4856	4744

11 I certify that the statements made by me above are correct and complete

Signature and Title of Editor, Publisher, Business Manager, or Owner
Michael J. Kelly, Managing Editor

PS Form 3526, January 1991 *(See instructions on reverse)*

HARVARD

The Ethics of Authenticity

Charles Taylor

In a discussion of ideas and ideologies from Gail Sheehy to Friedrich Nietzsche, from Allan Bloom to Michel Foucault, Taylor separates the good from the harmful in the modern cultivation of an authentic self. He sets forth the entire network of thought and morals that link our quest for self-creation with our impulse toward self-fashioning.

$17.95 cloth

Renewing Philosophy

Hilary Putnam

Hilary Putnam contests the view that only science offers an appropriate model for philosophical inquiry. His discussion of topics from artificial intelligence to natural selection, and of reductive philosophical views derived from these models, identifies the insuperable problems encountered when philosophy ignores the normative or attempts to reduce it to something else. Putnam takes up the problems posed by religious discourse and shows how the philosopher's frequently misunderstood forays into religious discourse actually open up philosophy to a broad range of practical, moral, and political issues

$25.00 cloth

Enlightenment, Revolution, and Romanticism

The Genesis of Modern German Political Thought, 1790–1800

Frederick C. Beiser

In a rich narrative history of ideas that proceeds from his highly acclaimed The Fate of Reason, Frederick Beiser shows how the French Revolution, with a rationalism and an irrationalism that altered the world, transformed and politicized German philosophy and its central concern: the authority and limits of reason.

34 halftones/$45.00 cloth

Truth and Objectivity

Crispin Wright

"This is original writing of a very high quality on some of the most fundamental issues in philosophy. It is the work of one who has a command of the many areas of philosophy required to address these general issues about truth. The writing is always clear . . . I would expect this book to become a major focus of discussion in contemporary debates about realism."

—Christopher Peacocke,
Magdalen College, Oxford

$29.95 cloth

Reasoning and the Logic of Things

Charles Sanders Peirce

Edited by Kenneth Laine Ketner

With an Introduction by Kenneth Laine Ketner and Hilary Putnam

With this publication, readers at long last are offered the philosopher's only known, complete, and coherent account of his own work. Comprising a series of lectures given in Cambridge, Massachusetts, in 1898, *Reasoning and the Logic of Things* is the most accessible and thorough introduction to Peirce's mature thought to be found within the compass of a single book.

10 halftones, 46 line illus.

$45.00 cloth/$22.50 paper

Quantum Mechanics and Experience

David Z. Albert

David Albert's lively account of the foundations of quantum mechanics is at once elementary and deeply challenging. It is an introduction accessible to anyone with high school mathematics and, at the same time, a rigorous discussion of the most important recent advances in our understanding of quantum physics, a number of them made by the author himself.

38 line illus.

$27.95 cloth

HARVARD

Inequality Reexamined

Amartya Sen

In this deft analysis, Amartya Sen argues that the dictum "all people are created equal" serves largely to deflect attention from the fact that we differ in age, gender, talents, and physical abilities, as well as in material advantages and social background. He argues for concentrating on higher and more basic values: individual capabilities and freedom to achieve objectives. By concentrating on the equity and efficiency of social arrangements in promoting freedoms and capabilities of individuals, Sen adds an important new angle to arguments about such vital issues as gender inequalities, welfare policies, affirmative action, and public provision of health care.

A Russell Sage Foundation Book

$29.95 cloth

NEW IN PAPER

Michel Foucault

Didier Eribon

Translated by Betsy Wing

"Didier Eribon has produced an astonishingly readable account of the man and his ideas, which is also in many ways an intellectual history of postwar France, so wide a trail did Foucault blaze in his time [A] sensitive, lucid and wide-ranging intellectual biography."

—Thomas Frick, *Los Angeles Times*

31 halftones/$14.95 paper

The Realm of Rights

Judith Jarvis Thomson

"This book isn't only about rights; it is also about thinking about rights. Thomson . . shows us, with elegance and wit, how to do this sort of work. what philosophical argument is, how one shapes an argument and makes it stick, and why the enterprise is so engaging" —Michael Walzer, Institute for Advanced Study, Princeton

$16.95 paper

ISBN: 0-674-74949-9

Pursuit of Truth

Revised Edition

W.V. Quine

"*Pursuit of Truth* is, like all Quine's works, most engagingly written. He strives for concision of the most telegraphic sort, and achieves it in a way that gives the special pleasure evoked by a conspicuous skill"

—Anthony Quinton, *The Times* (London)

$10.95 paper

Wise Choices, Apt Feelings

A Theory of Normative Judgment

Allan Gibbard

"A very distinguished book, rich in insights and written with a terse elegance that makes for wonderful clarity and ease of reading."

—John McDowell, University of Pittsburgh

$16.95 paper

Dialogues with Children

Gareth B. Matthews

Foreword by Robert Coles

Every week for a year, a professional philosopher and eight children at a school in Edinburgh met to craft stories reflecting philosophical problems. The philosopher, Gareth B Matthews, believes that children are far more able and eager to think abstractly than adults generally recognize. This engaging book has profound implications for education and for our understanding of the range of relationships between adults and children

$9.95 paper

Harvard University Press

Cambridge, MA 02138
(617) 495-2480

With BOOKS IN MIND

CONSCIOUSNESS RECONSIDERED

Owen Flanagan

"A joy to read, this book is grounded in common sense and never loses sight
of why consciousness is such a fascinating subject for study" — Edward M Hundert, M D , Harvard Medical School

A Bradford Book 276 pp , 7 illus $24 95

THE REDISCOVERY OF THE MIND

John R. Searle

"This book challenges the deepest assumptions of the 'cognitive science' movement
If Searle is right, we will have to revise our whole way of studying the mind Even if he isn't, this is sure to be among the most controversial philosophical publications of the decade"
— Ernie LePore, Rutgers University

Representation and Mind Series
A Bradford Book 304 pp $22 50

THE COMPUTATIONAL BRAIN

Patricia S. Churchland and Terrence J. Sejnowski

" A rewarding experience to read "
— J P Changeux, Institut Pasteur

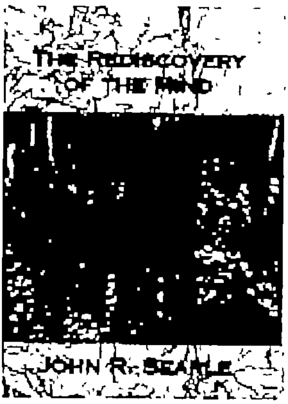
Computational Neuroscience Series 544 pp , 366 illus $39 95

SYMMETRY, CAUSALITY, MIND

Michael Leyton

"This is a superb books The writing is extraordinarily clear and the theory is very original It challenges the reader to think about basic questions and presents new and bold principles"
— Wayne Wickelgren, Columbia University

A Bradford Book 630 pp , 281 illus $45 00



Now Available For Text Use

THINKING THINGS THROUGH

An Introduction to Philosophical Issues and Achievements

Clark Glymour

"Glymour's book should make a good many people proud to be philosophers in the twentieth century"
— Douglas Stalker, University of Delaware

"…A unique kind of introductory textbook in philosophy, original in its conception of philosophy as a relatively formal subject encompassing mainly logic, probability theory, the theory of computation, and artificial intelligence This is a work that will do great good for the subject"
— Gilbert Harman, Princeton University

420 pp , 25 illus $35 00

In Paperback

LOGIC PRIMER

Colin Allen and Michael Hand

A Bradford Book 188 pp $12 95 paper

THEORIES OF TRUTH
A Critical Introduction

Richard L. Kirkham

"Kirkham has accomplished just what he set out to do, to write an accessible but accurate introduction to an area of philosophy that is a mystery even to many professional philosophers—theory of truth. To my knowledge, there is no other such introduction."
— Richard Foley, Rutgers University

408 pp $35.00

LANGUAGES OF THE MIND
Essays on Mental Representation

Ray Jackendoff

"It is in this collection that one can really appreciate the extent, breadth, creativity, and power of Jackendoff's work. It will be central to all of us concerned with progress in cognitive science."
— Barbara Landau, University of California, Irvine

243 pp., 31 illus. $25.00

FREUD'S DREAM
A Complete Interdisciplinary Science of Mind

Patricia Kitcher

By examining the rise and fall of psychoanalysis, Freud's Dream provides an extended case study of the appeal and potential dangers of the interdisciplinary approach to theory construction.

276 pp $24.95

WHAT IS COGNITIVE SCIENCE?

Barbara Von Eckardt

"This important books accomplishes a remarkable feat. It lucidly articulates and critically examines the shared assumptions that define the field of cognitive science at a time when many have come to suspect that cognitive science is a field in name only."
— Owen Flanagan, Wellesley College

440 pp $45.00 (December)

THE BOUNDS OF LOGIC
A Generalized Viewpoint

Gila Sher

A Bradford Book 178 pp $27.50

To order call toll-free 1 800 356 0343 or 617 628 8569. MasterCard & VISA accepted

A STUDY OF CONCEPTS

Christopher Peacocke

"Philosophy may not have a more important question than 'what are concepts?' Peacocke's answer is systematic philosophy in the grand tradition. It is also wonderfully clear, highly original, far-reaching in its consequences, and important."— Stephen Schiffer, The City University of New York

260 pp $29.95

LIAISONS
Philosophy Meets the Cognitive and Social Science

Alvin I. Goldman

"These essays have the qualities that distinguish Goldman's work: lucidity, rigor, broad knowledge, and pursuit of major issues."
— Philip Kitcher, University of California, San Diego

A Bradford Book 352 pp $39.95

CREATIVE COGNITION
Theory, Research, and Applications

Ronald A. Finke, Thomas B. Ward, and Steven M. Smith

240 pp 48 illus. $24.95

Now Available in Paperback

A THEORY OF CONTENT AND OTHER ESSAYS

Jerry A. Fodor

A Bradford Book 270 pp $14.95 paper

THE METAPHYSICS OF MEANING

Jerrold J. Katz

Representation and Mind Series
A Bradford Book 372 pp $15.95 paper

A NEW ERA IN COMPUTATION

edited by W. Daniel Hillis and James Bailey

A *Daedalus* special issue
250 pp $11.95 paper (December)

The MIT Press
55 Hayward Street, Cambridge, MA 02142

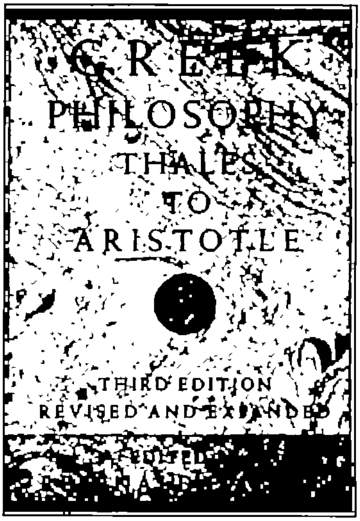
Classics in Philosophy

GREEK PHILOSOPHY
Thales to Aristotle

Third Edition, Revised and Expanded

Reginald E. Allen, Editor

Widely praised for its concentration on the issues that are most central to the history of Greek philosophy as well as for its accessibility, *Greek Philosophy: Thales to Aristotle* offers a valuable introduction to the works of the Presocratics, Plato, and Aristotle. Emphasizing metaphysics and epistemology as the foci of Greek philosophy, Allen's book also affords a solid basis for studying the majority of post-classical philosophers who have further explored metaphysics and epistemology as their central concerns.



1991 ISBN: 0-02-900495-0
$14.95 paper

Including key texts in Socratic ethics, Platonic politics, and natural science, the complete scope of Allen's collection has made it a perennial favorite, the most popular reader available on Greek philosophy. While the third edition retains the earlier emphasis on metaphysics and epistemology, it is enlarged to include more material on Greek science and mathematics, particularly with regards to analysis of infinity. Some of Aristotle's less well-known texts from his Physics have also been added. To complete the picture, revised translations of Plato's Symposium and Aristotle's Categories render these texts more faithful to the original Greek.

Reginald E. Allen is professor of classics and philosophy at Northwestern University.

THE FREE PRESS

POLITICAL IDEAS OF ST. THOMAS AQUINAS

St. Thomas Aquinas

0-02-840380-0 $10.95 paper

EIGHTEENTH-CENTURY PHILOSOPHY

Lewis W. Beck, Editor

0-02-902100-6 $12.95 paper

A MODERN INTRODUCTION TO PHILOSOPHY, 3rd Edition

Paul Edwards & Arthur Pap, Editors

0-02-909200-0 $27.95

A SHORT HISTORY OF CHINESE PHILOSOPHY

Fung Yu-Lan

0-02-910980-9 $14 95 paper

NINETEENTH-CENTURY PHILOSOPHY

Patrick Gardiner, Editor

0-02-911220-6 $12 95 paper

INTRODUCTORY READINGS IN AESTHETICS

John Hospers

0-02-915260-7 $14.95 paper

MORAL AND POLITICAL PHILOSOPHY

David Hume

0-02-846170-3 $12.95 paper

DIALOGUES CONCERNING NATURAL RELIGION

David Hume

0-02-846180-0 $10.95 paper

PHILOSOPHY OF THE SIXTEENTH AND SEVENTEENTH CENTURIES

Richard H. Popkin

0-02-925490-6 $12.95 paper

ESSAYS IN PRAGMATISM

William James

0-02-847140-7 $9.95 paper

CRITIQUE OF JUDGEMENT

Immanuel Kant

0-02-847500-3 $12.95 paper

PHILOSOPHY AND TECHNOLOGY

Carl Mitcham & Robert Mackey, Editors

0-02-921430-0 $14.95 paper

THE SPIRIT OF THE LAWS

Baron de Montesquieu

0-02-849270-6 $13.95 paper

NEWTON'S PHILOSOPHY OF NATURE

Isaac Newton

0-02-849700-7 $12.95 paper

A CRITICAL HISTORY OF WESTERN PHILOSOPHY

D.J. O'Connor

0-02-923840-4 $14.95 paper

GREEK AND ROMAN PHILOSOPHY AFTER ARISTOTLE

Jason L. Saunders

0-02-927730-2 $14.95 paper

ETHICS

Benedict de Spinoza

0-02-832630-3 $10.95 paper

TWENTIETH CENTURY PHILOSOPHY

The Analytic Tradition

Morris Weitz

0-02-934990-7 $12.95 paper

MEDIEVAL PHILOSOPHY

John F. Wippel & Allan B. Wolter

0-02-935650-4 $12.95 paper

Attention College Professors:

Order your examination copy today! Simply place your request on university stationery, indicating course name, current text, projected enrollment, and the date your next text decision is due. Send to.

Fp **THE FREE PRESS**

A Division of Macmillan, Inc./A Maxwell Macmillan Company

866 Third Avenue, New York, NY 10022

PHILOSOPHY

THE NIETZSCHE LEGACY IN GERMANY, 1890-1990
STEVEN E. ASCHHEIM
"One of the most important works of German and European intellectual history published in years. . . . A long overdue history of the multivalent reception and reworking of Nietzsche."
—Jeffrey Herf,
author of *Reactionary Modernism*
Weimar & Now: German Cultural Criticism, $40.00 cloth, illustrated

INFERENCE, EXPLANATION, AND OTHER FRUSTRATIONS
Essays in the Philosophy of Science
Edited by JOHN EARMAN
These provocative essays illuminate the uncertainty and excitement in this changing field.
Pittsburgh Series in Philosophy and History of Science
$15 00 paper, $47 50 cloth

THE MORAL COMMONWEALTH
Social Theory and the Promise of Community
PHILIP SELZNICK
"Outstanding. . . . A book no student of our social or moral foundation, condition, and hence future can afford to miss." —Amitai Etzioni,
author of *The Moral Dimension*
A Centennial Book, $40.00 cloth

PLATO'S EUTHYDEMUS
Analysis of What Is and Is Not Philosophy
THOMAS H. CHANCE
"An indispensable tool for penetrating to the sources of Plato's thinking"
—G. B. Kerferd, University of Manchester, *$40.00 cloth*

APOCALYPSE AND/OR METAMORPHOSIS
NORMAN O. BROWN
New in paper—"Brown employs ideas in such a volatile manner that every reading and re-reading is an act of creation. A book for the ages."
—Washington International Arts Letter
$12.00 paper

POLITICAL CRITICISM
IAN SHAPIRO
New in paper—"A provocative and compelling vision of political theory."— *Dissent*
$15.00 paper

At bookstores or order
1-800-822-6657.

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY LOS ANGELES NEW YORK

Arendt, Camus, and Modern Rebellion

Jeffrey C. Isaac

"Isaac draws together Arendt and Camus into a fruitful dialogue. That dialogue is wide-ranging and tackles major events of the twentieth century, the creation and terrible denouement of totalitarianism, and, in its aftermath, the possibility for democratic politics, the forging of human identity, and the sustaining of anything like an ethical life. The work is very important."—Jean Bethke Elshtain $30.00

Hermeneutics Ancient and Modern

Gerald L. Bruns

Bruns expands our sense of the horizons of hermeneutics by situating its basic questions against a background of philosophical thought from Plato to Ricoeur to Rorty. $37.50

The Future of Liberal Revolution

Bruce A. Ackerman

An eminent political philosopher discusses how liberal democracy can be achieved in Western and Eastern Europe. $18.50

Yale University Press

92A Yale Station
New Haven, CT 06520

The Dialogue of Justice

Toward a Self-Reflective Society

James S. Fishkin

In his latest book, James Fishkin offers a theory of liberal democracy based on a self-reflective society—a participatory democracy that would provide the key to issues of political legitimacy and social justice.

"A tour de force. A display of argumentative virtuosity of a high order."—John Gray $27.50

New in Paperbound

Ethics in the Sanctuary

Examining the Practices of Organized Religion

Margaret P. Battin

In the first book of its kind, philosopher Margaret P. Battin explores a compelling range of ethical dilemmas raised by organized religion.

"This is a book which pioneers a new area in professional ethics."—Peter Sedgwick, *Journal of Applied Philosophy* $16.00

The Faces of Injustice

Judith N. Shklar

A distinguished political theorist formulates a new political and moral theory of injustice that encompasses not only deliberate acts of cruelty or unfairness but also indifference to such acts.

"A provocative meditation on the meaning and the scope of human injustice."—Martha Nussbaum, *New Republic* $10.00

The Storrs Lectures on Jurisprudence

OUTSTANDING PHILOSOPHY FROM CAMBRIDGE

Privacy and Social Freedom

Ferdinand David Schoeman

This book attacks the assumption found in much moral philosophy that social control as such is an intellectually and morally destructive force. It replaces this view with a richer and deeper perspective on the nature of social character aimed at showing how social freedom cannot mean immunity from social pressure

Cambridge Studies in Philosophy and Public Policy
41564-0 Hardcover $49.95

The Ethics of Trade and Aid

U.S. Food Policy, Foreign Competition, and the Social Contract

Paul B. Thompson

Professor Thompson replaces the traditional military-territorial model of the nation state with the notion of the trading state. What emerges is a new interpretation of social contract theory that can determine goals for international trade and development policy.

Cambridge Studies in Philosophy and Public Policy
41468-7 Hardcover $49 95

Drugs and Rights

Douglas N. Husak

Do adults have the right to use drugs for recreational purposes? The first serious work to address the question argues that the "war on drugs" violates the rights of adults wanting to use drugs for pleasure, and that criminal laws against this use are incompatible with moral rights.

Cambridge Studies in Philosophy and Public Policy
41739-2 Hardcover $49 95
42727-4 Paper $17 95

Economic Rights

Ellen Frankel Paul,
Fred Miller, Jr.,
Jeffrey Paul, Editors

These essays explore the kinds of economic rights people should have, how they should be conceived and grounded, and the consequences that would ensue were one system or another set up to ensure them.

Contributors *N Scott Arnold, Jan Narveson, Randy E. Barnett, Daniel M. Hausman, James W Child, Jonathan R. Macey, Eric Mack, Richard J. Arneson, David Copp, Gregory S. Kavka, John E. Roemer.*

Social Philosophy and Policy 9:1
42873-4 Paper $19 95

The Good Life and the Human Good

Ellen Frankel Paul,
Fred Miller, Jr.,
Jeffrey Paul, Editors

What is the good life? This question captured the attention of ancient philosophers, and it remains with us today because it compels us to consider what it is to be human.

Contributors: *Julia Annas, Lawrence C Becker, James Griffin, Thomas Hurka, Shelly Kagan, John Kekes, Frank I. Michelman, Warren Quinn, Amelie Oksenberg Rorty, L W. Sumner, Judith Jarvis Thomson*

Social Philosophy and Policy 9:2
43759-8 Paper $19 95

Available in bookstores or from

CAMBRIDGE
UNIVERSITY PRESS

40 W. 20th St., N.Y., NY 10011-4211
Call toll-free 800-872-7423
MasterCard/VISA accepted
Prices subject to change